RTO-MP-4
AC/323(HFM)TP/2

RTO-MP-4

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION
BP 25, 7 RUE ANCELLE, F-92201 NEUILLY-SUR-SEINE CEDEX, FRANCE

RTO MEETING PROCEEDINGS 4

Collaborative Crew Performance in Complex Operational Systems

(l'Efficacité du travail en équipage dans des systèmes opérationnels complexes)

Papers presented at the RTO Human Factors and Medicine Panel (HFM) Symposium held in Edinburgh, United Kingdom, 20-22 April 1998.

19990902 014

Published December 1998

Distribution and Availability on Back Cover

RTO-MP-4
AC/323(HFM)TP/2

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION

BP 25, 7 RUE ANCELLE, F-92201 NEUILLY-SUR-SEINE CEDEX, FRANCE

RTO MEETING PROCEEDINGS 4

Collaborative Crew Performance in Complex Operational Systems

(l'Efficacité du travail en équipage dans des systèmes opérationnels complexes)

Papers presented at the RTO Human Factors and Medicine Panel (HFM) Symposium held in Edinburgh, United Kingdom, 20-22 April 1998.

The Research and Technology Organization (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote cooperative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective coordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also coordinates RTO's cooperation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of initial cooperation.

The total spectrum of R&T activities is covered by 6 Panels, dealing with:

- SAS Studies, Analysis and Simulation
- SCI Systems Concepts and Integration
- SET Sensors and Electronics Technology
- IST Information Systems Technology
- AVT Applied Vehicle Technology
- HFM Human Factors and Medicine

These Panels are made up of national representatives as well as generally recognised 'world class' scientists. The Panels also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier cooperation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

The content of this publication has been reproduced
directly from material supplied by RTO or the authors.

Printed on recycled paper

Published December 1998

ISBN 92-837-1008-8

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

Collaborative Crew Performance in Complex Operational Systems

(RTO MP-4)

Executive Summary

The first symposium of the Human Factors and Medical Panel of the Research and Technology Organization (RTO) focused on the theme *collaborative crew performance in complex operational systems*. This theme denotes two primary trends that are applicable to systems research, design, and engineering: a) collaboration and b) complexity. As new systems are planned and fielded in the next millennium, collaborative crew performance in complex operations will need to be addressed in effective ways to enable successful missions. This symposium's goal has been to identify the multi-dimensional problems and challenges that scientists and engineers encounter as they consider crews in complex systems; and then take a global look at innovative solutions. The problems, challenges, and solutions generated span across many countries; represent a plethora of philosophies, theories, frameworks, approaches, designs, technologies, measures, and contexts; and provide a cornerstone for understanding the many constraints inherent in crew activities. The papers represent an outstanding gathering of experts, specialists, and researchers in this field and are cogent for answering questions such as:

- How to establish research programs in collaborative crew performance?
- What are the influences of individual differences in crew operations?
- How are crew measurements different from individual measures?
- How can cognitive engineering be applied to the design of collaborative systems?
- What theories of groups are relevant to study crews in complex domains?
- How would an ethnographic approach to crew interface design be implemented?
- What collaborative technologies are available to enhance crew performance?
- What collaborative paradigms / tasks are available to researchers?
- What real world, collaborative domains have been studied in-depth?
- What are some of the socio-cultural factors that affect crew performance?
- How can collaborative task analysis be useful for complex systems design?

These questions (and many more) can be answered by perusing author's papers. In this sense, the proceedings provide a kind of exploratorium for those who will come into contact with the themes of collaboration and complexity. The intent has been to provide a nexus of thought and direction with respect to *collaborative crew performance in complex operational systems*.

L'efficacité du travail en équipage dans des systèmes opérationnels complexes

(RTO MP-4)

Synthèse

Le premier symposium organisé par la Commission facteurs humains et médecine de l'Organisation pour la recherche et la technologie (RTO) a eu pour thème *l'efficacité du travail en équipage dans des systèmes opérationnels complexes*. Ce thème dénote deux grandes tendances communes à la recherche, à la conception et à l'ingénierie des systèmes à savoir: a) la collaboration et b) la complexité.

Avec la réalisation et la mise en service de nouveaux systèmes au prochain millénaire, les différents aspects de cette question devront être traités de façon constructive afin d'assurer la réussite des missions. Ce symposium a eu pour objectif d'identifier d'abord les problèmes multidimensionnels et les défis auxquels les scientifiques et ingénieurs sont confrontés lorsqu'il s'agit d'intégrer des équipages dans des systèmes complexes; et ensuite de procéder à l'examen de l'ensemble des solutions novatrices proposées.

Les problèmes, défis et solutions dégagées, qui concernent de nombreux pays, ont fait l'objet d'un nombre pléthorique de théories, de philosophies, de technologies, d'études, d'approches, de contextes, et de mesures. Ils sont la pierre angulaire de la connaissance des nombreuses contraintes propres aux activités des équipages. Les communications, la contribution d'un rassemblement d'éminents spécialistes et chercheurs dans ce domaine, permettent de répondre à un certain nombre de questions telles que:

- comment établir des programmes de recherche en matière de travail collectif en équipage?
- quelles sont les influences des différences des individus lors des opérations en équipage?
- dans quelle mesure les données équipage sont-elles différentes des données individuelles?
- quelles sont les applications de l'ingénierie cognitive pour la conception de systèmes coopératifs?
- quelles sont les théories de groupes applicables à l'étude du comportement des équipages en environnement complexe?
- comment mettre en oeuvre une approche ethnographique de la conception des interfaces homme/machine?
- quelles sont les technologies coopératives proposées pour l'amélioration des performances des équipages?
- quels sont les paradigmes/tâches coopératifs à la disposition des chercheurs?
- quels sont les domaines coopératifs concrets ayant fait l'objet d'études approfondies?
- quels sont les principaux facteurs socioculturels agissant sur les performances des équipages?
- l'analyse coopérative des tâches, peut-elle servir à la conception de systèmes complexes?

Une lecture attentive des communications du symposium fournira les réponses à ces questions et à bien d'autres encore. Ainsi, le compte rendu de la réunion est en quelque sorte un *exploratorium* pour tous ceux concernés par les sujets du travail en équipage et de la complexité. L'idée directrice du symposium a été de présenter une série de réflexions sur *l'efficacité du travail en équipage dans des systèmes opérationnels complexes.*

Contents

Preface

Specialists within the human factors and medical community are now frequently coming into contact with real world settings that demand collaborative crew interaction. These settings are complex, contain distributed - highly technological - systems, and require joint interdependence to establish meaningful operations. Yet the research and design data to support these demands has not been well developed, is poorly organized, and is typically not generated for complex operational domains. Given these states, the purpose of this symposium was to congeal and organize topics that address both the theoretical and practical concerns of those specialists who are now engaged with some element of collaborative crew performance in complex operational systems. Hence, these proceedings enact a repository for those currently addressing this area and for those who will advance this area in the future. The topics within are wide-ranging and should be edifying for differing quests for knowledge in this area. It is the hope of the organizers that this symposium has laid down (1) a foundation for potential contributions to collaborative crew performance and (2) a spirit of congenial aspiration that will sustain research and development to new levels of maturity and advancement.

Human Factors and Medicine Panel

Chairperson: Dr M.C. WALKER
Director, Centre for Human Sciences
F138 Building - Room 204
DERA
Farnborough, Hants GU14 0LX, U.K.

TECHNICAL PROGRAMME COMMITTEE

Co-Chairpersons

Dr K.BOFF
AFRL/HE
2610 Seventh Street
Wright-Patterson, AFB, Oh 45433-7901
USA

Telephone: (1) 937 255 6327/6328
Telefax: (1) 937 656 7617
Email: kboff@al.wpafb.af.mil

Ms. J. DAVIES
British Aerospace Defence Ltd.
Aircraft Division
Farnborough Aerospace Centre
P O Box 87
Farnborough, Hants GU14 6YU, UK
Telephone: (44) 1252 382 282
Telefax: (44) 1252 382 060
Email: jo.davies@bae.co.uk

Members

Cdr. D.L. DOLGIN, MC, USN
Crew Systems Department
Code 4.6M - Bldg 2187, Ste 1282
Naval Air Warfare Ctr. - Aircraft Division
48110 Shaw Road, Unit #5
Patuxent River, MD 20670-1906
Telephone: (1) 703 696 4265
Telefax: (1) 703 696 1245
Email: dolgindl.ntrprs@navair.navy.mil

Dr G.M. ROOD
Manager Mission Management Dept.
Defence Research Agency
Room 204, Bldg R177
Farnborough, Hants GU14 6TD
United Kingdom
Telephone: (44) 1252 393076
Telefax: (44) 1252 393091
Email: gmrood@dra.hmg.gb

Dr A. LEGER
SEXTANT AVIONIQUE
Rue Toussaint Castros
BP 91
33166 St Médard en Jalles cedex
France
Telephone: (33) (0)5 56 13 51 68
Telefax: (33) (0)5 56 13 50 37
Email: alain.leger@sextant.thomson.fr

Prof. dr. D. VAN NORREN
Aeromedical Medical Institute
Kampweg 3
Postbus 22
3769 ZG Soesterberg
The Netherlands
Telephone: (31) 346 334 368
Telefax: (31) 346 351 899
Email: d.vannorren@nlrgc.antenna.nl

LOCAL COORDINATOR

Dr A.J.F. MACMILLAN
Royal Air Force School of Aviation Medicine
Farnborough, Hants, GU14 6TD, U. K.
Telephone: (44) 1252 392 635
Fax: (44) 1252 393469
Email: central-sam@btinternet.com

PANEL EXECUTIVE

From Europe and Canada:
Dr C. WIENTJES
RTA/NATO
BP 25, 7, Rue Ancelle
F-92201 Neuilly-sur-Seine, Cedex, France

From USA
RTA/NATO/HFM
PSC 116
APO AE 09777

Tel.: (Paris) 1 55 61 22 60/62 - Telex: 610176F
Telefax: 1 55 61 22 99/98
Email: wientjesc@rta.nato.int

COLLABORATIVE CREW PERFORMANCE IN COMPLEX OPERATIONAL SYSTEMS: A FINAL SUMMARY

M. D. McNeese
United States Air Force Research Laboratory
Collaborative Systems Technology Branch
AFRL / HECI, 2255 H. Street, Bldg 248
Wright-Patterson Air Force Base, OH USA 45433-7022

SUMMARY

The technical summary of this RTO symposium is provided. The reader is given an integrative structure that comprehensively describes the content and directions of six distinct elements of research within collaborative crew performance. Each element is individually reviewed to construct a common ground for the prospective reader. Issues, challenges, examples, and trends are identified to assess the topics that have been presented. Concluding remarks attempt to make rhyme and reason out of the levels of diversity experienced while honoring the principle of mutual learning in addressing the difficulties of convergence.

INTRODUCTION

Orientation of the Symposium

The first symposium of the RTO Human Factors and Medicine Panel has focused on the topic of Collaborative Crew Performance in Complex Operational Systems. Collaboration among crew members, and their associated human system interface requirements, are increasingly becoming a critical factor for enhancing readiness, exacting coordinated actions, and generating new levels of shared situational awareness within a variety of situated operational contexts. Resident within collaborative crew performance is the reliance upon computers, communications, and cognitive interactions among various factions. Often this reliance is directly impacted by (1) the human systems interface designed to augment collocated and distributed collaboration (2) the human factor's analysis tools (e.g., cognitive engineering) used to distill collaborative knowledge and design needs and (3) knowledge about the crew's context itself (e.g., specific affordances of crews in long-duration flights). It is not just sufficient to 'adapt' single user models, individualistic analytical tools, or unitary views of situational awareness for collaborative crew orientation. What is required is to address *Collaborative Crew Performance in Complex Operational Systems* on its own terms from the first stage of conceptualization; and to continue to uniquely develop appropriate models, tools, approaches, interfaces, and simulations with the relevant focus on understanding crew inter-activity, crew support systems, and crew environments. Hence, the purpose of this symposium has been to establish and promulgate a broad bandwidth of new research and development approaches to complex operational systems distinctively from a 'crew-centered' perspective. The intent has been to show the potential contributions from the confluence of many interdisciplinary research thrusts (e.g., human-computer interaction, computer-supported cooperative work, naturalistic decision making, multi-crew aiding, cognitive engineering, situated cognition, cooperative learning) designed to highlight new advances and critical research agendas in collaborative crew performance.

Role of the Recorder

My role in the symposium has been to record and summarize the presentations and provide discrimination of meaningful tones of the meeting. This requires integration of themes and thrusts, reviewing state-of-the-art technologies that are presented, and establishing an understanding of the research agendas that scientists propose in their talks. By knitting together these areas (and others) and through an assessment of the various topics presented, an overall summary of the symposia can begin to unfold. Inherently, this should include various challenges and issues that will need to be addressed. The role of the recorder also requires that one sense what has been done in the past (historical precedence) while becoming aware of what scientists are proposing as the next level of advancement (the future). As part of the recorder role, it is also necessary to report the different applications and operational domains in which crew collaboration occurs.

I have approached the role of recorder with two main assumptions in mind. First, I have engaged the meeting from an ethnographic perspective. That is, many notes where taken during the symposium during the course of the presentations which are supplemented by audio recordings and the papers themselves. Although the intent was to be an objective observer of the event and to utilize various tracings of the event to derive levels of understanding, other more subjective interpretations of the meeting ensued as well, mainly in the form of conversations with different scientists and presenters. Together these various sources of input have provided the necessary fabric that defines the "recording" of the symposium. The second assumption is that given these sources of input it is not possible to report every nuance and detail captured during the recording process. Therein, a strategy for reporting and documenting the symposium in an efficient manner is needed.

The strategy I have utilized is one of structure and priority. The structure for the written summary enables potential readers the opportunity to sample a broad cross-section of expertise on collaborative crew performance. Priority is used to report the level of depth of knowledge. Because the complete level of depth is not possible in this space, only selected items within certain structural components will be examined in depth. Priority is also maintained by emphasizing those areas that tend to appear repeatedly

across papers. Priority of report can thus be determined as a function of consensus based on examining what presenters have done in addressing collaborative crew performance. In addressing priorities (via the categorical structure) it necessarily means exercising *selective attention* to ideas in one paper perhaps at the exclusion of the same point made in another paper. Some points that should have been made may fall through the crack. Selective attention thus means that as recorder I am utilizing points or papers that are representative of the objective I have in mind. Finally, this summary is wholly created, limited, and structured within the confines of my personal observations and interactions as filtered through experience, biases, and ideation. Another ethnographer might have a new take on the thoughts presented, therein the "recordings" are intrinsic to my purview. With these assumptions and rules of engagement now extant, it is now time to delve into the content of this gathering.

CATEGORICAL STRUCTURE OF THE EVENT

Dr. Boff introduced the symposium with a collage that visually expressed the overall essence of *Collaborative Crew Performance in Complex Operational Systems.* Many of us have heard or been associated with terms like collaboration or performance or systems or operational. Certainly, the collage is representative of these descriptors. But, clearly one of the most important and replicated aspects of this conference is the theme of *complexity.* Complexity may be communicated through various expressions as was evident in many of the papers. How one addresses complexity through each organizing element of the categorical structure affects our understanding of collaborative performance in operational systems. In this sense, I believe that complexity is one of the defining characteristics of the many papers presented. It is instructive to keep the cornerstone of complexity at the heart and center of the structure defined to help summarize this symposium.

As mentioned in the last section, structure and priorities are necessary ingredients to properly entwine the multiple dimensions of the symposium into an integrated whole. Figure 1 provides the categorical structure selected to record the symposium. As one can see there are six main elements that will be addressed during the remainder of this summary. These elements are placeholders for various units of information that appear as part of the symposium and in that sense represent "organizing elements" that are central in portraying both breadth and depth. I have stratified all the papers using this scheme and each element contains roughly the same amount of papers. The symposium papers seem to be distributed rather equally using this organizational scheme.

The organizing elements are crucial not only because they interrelate and entwine the multiple dimensions of collaborative crew performance and operational systems but because they anchor the emergence of overriding trends. Refer to Figure 2. The trends to watch for are: (1) *research issues* (2) *design directions* (3) and *future progress and agendas.* In many ways these trends are still in formation and undergoing multiple changes. This indicates that the areas of concern within this symposium are still evolving, growing, and maturing. Hence, there is room for a plethora of viewpoints, approaches, issues, and design positions to coexist simultaneously. Indeed, this statement is reinforced by looking at the diversity within any organizing element. The remainder of the paper in fact will make an attempt to look at this diversity and make sense of it.

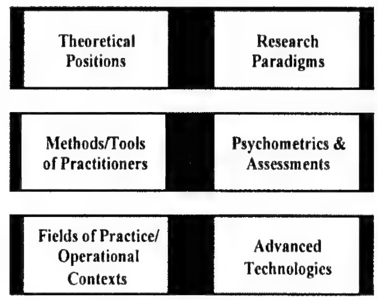


Figure 1. Six Elements of the Categorical Structure

Collaborative Crew Performance in Complex Operational Systems

Figure 2. Symposium Connections and Trends

THEORETICAL POSITIONS

One of the most salient issues (and challenges) in the symposium was how and where to start doing research and whether that research was meaningful in terms of what it informed. Unfortunately, this often took the form of a dichotomous breakdown between basic research / theoretical positions and field studies / contextual inquiries. In research and design programs this issue is sometimes manifest as *where one begins* their research for knowledge. For some this means starting in the research literature and reviewing theoretical positions which in turn gauges a top-down approach. For others it means starting with the context or field of practice and eliciting knowledge from experts and building up to experimental studies. However, as some incisive researchers observed this does not have to be an either-or choice.

Many of the presentations provided a series of experiments that were positions on more theoretical ground and therein tied to specific theoretical positions. This can be seen in the work of Linda Elliott and her associates at the U.S. Air Force Research Laboratory at

Brooks Air Force Base (Elliott, Schiflett, and Dalrymple; Cobb, Mathieu, Elliott, and Dalrymple in this volume).

This can be seen as Airborne Warning and Control System (AWACS) crews were evaluated using the theoretical underpinnings of Hackman's Input-Process-Output model. The trend in using theoretical positions to discover findings in crew performance is one in which the top-level theory (or model) predicates the entire approach to research through a series of mutually constrained conditions. For example, the Hackman team process model when applied to the AWACS team problem derives levels of independent variables, an experimental design, use of a scenario with decision events, specific measures of effectiveness, application of statistical analysis, model verification via explanation of results obtained, and finally utilizing results for specific applications (e.g., training, interface design, and technology needs). Although many other factors emerge from theoretical positions these are certainly typical and can be used to demonstrate the value of expanding knowledge in crew performance. It must be noted that the Elliott group is not solely using the theoretical approach as they also work closely subject matter experts and real world operations of AWACS crews. This is an example of blending the theoretical positions with other elements to provide a wholistic approach to research.

Another major research program that preserves theoretical concerns was given by one of the keynote speakers, Joan Johnston, Naval Air Warfare Center Training Systems Division. The program incorporated a wholistic view by strongly interconnecting theoretical positions with training strategies and operational needs. In their paper, Johnston, Cannon-Bowers, and Salas (this volume) indicate the desire to formulate realistic team training tasks for a ship's Combat Information Center. Although their research agenda and tasks are highly informed from real world parameters, the research emanates from decision theory explored for many years in the *tactical decision making under stress initiative*. One other example from the Navy is provided in area of team situational awareness (Muñiz, Stout, Bowers, & Salas, this volume). In this example, team situational awareness is derived from SA theory and definitions. The team SA views are developed in the context of measurements for teams with high or low SA. The Navy group then developed the SALIANT methodology from these foundations and in turn create pertinent scenarios and studies. Another paper at the symposium (Bowers, Weaver, Barnett, & Stout, this volume) provides the empirical validation of the SALIANT methodology. With an approach that counts on theory, one can see the consequential focus on scenario development, simulator evaluations, and use of appropriate experimental measures as themes in the test of theory and real world problems. Of special note was Eric Hollnagel's invited paper - which even though it did not specifically endorse theoretical positions - noted some of the problems involved between modeling and measuring joint cognitive systems. Obviously, theories without measures (and intrinsic models) of phenomena are empty sets. As Hollnagel suggests the interpretation of measures is contingent on the degree of articulation extant within a model.

Some of the more salient issues raised were (1) whether theories (as a generalized set of principles) can successfully be transformed into ecologically valid positions suitable for use in discovering / evaluating real world problems (2) knowing what is the "right" theory to apply to crew situations given that some theories focus on team situational awareness or communication where others might focus on additional constructs such as team schema similarity or individual differences in teamwork (3) whether future agendas should focus more on developing new team theories to encompass integration of additional variables or to adapt existing theories with feedback from validations. Many issues resonate with the advantages / disadvantages with theory in general and were not solely specific to crew performance per se.

In sampling across the various papers the following theoretical (or quasi-theoretical) orientations were in evidence or mentioned as specific core values for researchers: shared mental models, team self correction, ecological psychology, team situational awareness, team compatibility, 'team of teams', social ergonomics, affordances, activity theory, actor-network theory, articulation work, folk versus articulated models, orderliness of a system, socio-technical theory, team member schema similarity, team effectiveness model, perceptual control theory, knowledges, distributed cognition, and the performance hierarchy. It is worth mentioning that the conceptual foundations established by Jens Rasmussen [1, 2] were specified in several papers by researchers. This is noteworthy for a few reasons. First, it suggests that Rasmussen's skill-rule-knowledge framework and/or the decision ladder model are actively being applied to crew / teamwork -level applications as opposed to just single operator domains. Second, these conceptualizations although originally distilled from ecological inquiry of a domain are now being used to guide and anchor both theoretical and field of practice research. Third, the framework was cited more frequently than any other across all presentations and thereby shows favorably for various goals and objectives. Fourth, the framework provides one basis for integrating research with design as *ecological interface design* concepts are beginning to be used by practitioners of human factors and medicine (refer to the paper by Beevis, Vicente, and Dinadis, this volume). The paper by Eggleston (this volume) provides a nice review of some of the *complexity issues and challenges* that span across research, cognitive engineering, and design.

For additional information on theoretical positions, I would specifically refer readers to the excellent papers by (1) Andre, Klinger, & Williges (this volume) that reviews theories and measures that are useful in studying the design and use of computer-augmented, distributed team decision making (2) Nosek (this volume) which reviews theoretical foundations for group sensemaking and the social construction of knowledge with specific focus on augmenting computer support (3) Cooke, Elder, & Ward (this volume) which elaborates a comprehensive look at the role of communication in aeronautical decision making (4) the keynote address by Liam Bannon (this volume) outlines many of intricacies between theory, work, and practice.

RESEARCH PARADIGMS

I have already alluded to several programs which evolved paradigms from theoretical positions. Many more research paradigms and studies where reported during the course of presentations. There is a two way street that can lead to the development of research paradigms and consequently the conduct of an experiment. One street proceeds from the top terrain - the decline from the top towards the bottom (as previously elaborated with several groups who have tightly coupled their paradigms with theoretical positions). Alternatively, the other street proceeds from the bottom upwards - the incline. The incline drives researchers to derive synthetic task environments from what they learn as ethnographers working in the context and with experts. As mentioned, the incline and decline are not mutually exclusive approaches as in several papers scientists merge the two roads as an effective by-pass to traverse complex conditions. Lets first take a look at some cross-sections that are indicative of current research paradigms in collaborative crew performance.

Rasker, Schraagen, Post, & Koster (this volume) supply us with a highly integrated research paradigm developed to study information exchange in command and control teams. They begin with an outline of a model of command and control functions and note the relation of this to Rasmussen's rule-based behavior in his decision ladder notation (1983). Hence, they conceptually take the high road to anchor their work. Note however that they also specify meaning for their paradigm by tracking field studies on information exchange in teams; specifically tailoring their work to team self-correction. The approach then develops research questions, develops the fire fighting paradigm to emulate command and control activities, and tests the various hypotheses on information exchange under differing communication conditions.

In contrast, the work of Callister, Percival, Retzlaff, and King (this volume) explores development of a research paradigm directly connected and immersed within a real world environment. Their research studied the effects of stress on individual and group problem solving and utilized USAF aircrews performing required survival, evasion, resistance, and escape training. Because this training takes place under extreme conditions within real world constraints it provides a unique opportunity to create a research paradigm to study stress effects. This also brings up the issue of whether "simulated conditions of stress" as enacted in laboratory studies are realistic enough to capture the stress as experienced in practice. In this research the focus is on measurement / assessment of cognitive performance and fatigue across different training regimens. Like so many of the paradigms, there is a distinct relation between top-down and bottom-up sources of knowledge. Here the measures assessing cognitive performance track upward to the constructs purported to be resident in cognitive theory (e.g., spatial memory, problem solving, vigilance). To the extend these individual measures can glom together for group-based objectives (command and control, situational awareness, and communications), the research paradigm spans both high and low roads.

The distributed cooperative planning work of Patricia Jones at the University of Illinois (this volume) is exemplary of the successful merging of the decline and incline. From the high road it is conceptually predicated from three aspects of CSCW: shared information spaces, articulation work, and social presence. From these aspects, Jones derives issues on how to share information, coordinate activity, and manage actors in collaborative spaces. From the low road, collaborative planning is developed from the context of the Army research project CoRAVEN and LorRaine Duffy's C2MUVE (this volume) for the Navy. Research derives from the study of the use of collaborative support interfaces in these projects. Such interfaces are prototyped from knowledge taken from the top level concepts in CSCW.

Some of the prime issues forthcoming from this organizing element are (1) knowing how to operationalize team-theoretic constructs as meaningful variables (2) validly determining measurement of these variables especially in terms of the team component of performance (3) how to simulate event or decision driven scenarios to capture complexity and realism (4) deciding on multiply-related tradeoffs such as cognitive fidelity, level of team expertise (i.e., real world experts or college-students as Ss), and team longevity necessitated for an experiment (5) how the design of experiments and the paradigm itself could be useful to the generation of crew support systems / crew interface designs (i.e., translation of research results into meaningful design criteria).

Several *research paradigms / simulation environments* were reported by scientists to include but not limited to the following cross-section sample: Decision Making Evaluation Facility (DEFTT), Situational Awareness Linked Instances Adapted to Novel Tasks (SALIANT), Communications Simulation Training and Research Systems (C3STARS), Event Based Approach to Training (EBAT), Scud Hunting Engagement Paradigm (SHEP), Concept of Operations Exercise (COOPEX), Team Training Paradigm, TORNADO Simulator, Fire Fighting Task, and Space Orbital Mission Simulation.

Of note for more in-depth understanding for many of the issues involved in establishing team research and research paradigms was the insightful presentation and paper by Johnston, Cannon-Bowers, and Salas (this volume). The paper by Andre, Kleiner, and Williges establishes a great foundation for developing theory, paradigms, measures, and verification for team decision making and communication. I must also note that this is one area that I think is now potentially ripe for more development; mainly to creatively introduce new elements of complexity that have been previously ignored. Emotions, group trust, belief systems, and team-intelligent system integration, and realistic stress in extreme environments are just some of the new challenges for research paradigms to incorporate.

METHODS / TOOLS OF THE PRACTITIONERS

Many methods, techniques, and tools of crew performance research are highly predicated from either (1) the theoretical positions or research paradigms a scientist uses or (2) the field of practice which one is studying. Therein this is a broad category which contains manifold approaches. However, one theme that resonated throughout papers within this area is that of cognitive engineering. The paper provided by Eggleston (this volume) thoughtfully describes and outlines issues associated with integrating cognitive engineering into design practice. He looks at cognitive engineering as a design method / technology capable of addressing weaknesses in the design of complex systems. He rightfully identifies gaps between the use of cognitive engineering methods in the crew station design process in addition to pointing out the many perceptions of cognitive engineering by systems managers and design practitioners. Eggleston's concluding point is worthy of consideration for collaborative crew performance as derived from cognitive engineering methods. It is that cognitive engineering can be considered progressive (rather than a passing fad) if the field can demonstrate contributions to design and be integrated into the larger system design process.

The work of Jens Rasmussen in utilizing cognitive engineering (as both a basis for analytical methods / tools and development of design artifacts) was dully noted by various contributors at this symposium. In particular, the decision ladder and abstraction hierarchy have been used as a basis to initiate ecological interface design. The paper by Beevis, Vicente, and Dinadis (this volume) is an excellent example of how ecological interface design methods have been used to develop interfaces in the CC-130 Hercules aircraft. This project orients cognitive engineering within an overall human engineering model for program development for advanced aircrew systems and as such may be useful in assuaging some of the concerns in the Eggleston paper. The authors make use of the following methods in a systematic way: protocol analysis, preparation of an abstraction hierarchy, the s-r-k decision ladder as a basis for defining levels of information support, and a rapid prototype of an interface which was subsequently evaluated by flight engineers. This is one of the first applications of ecological interface design to crew applications (rather than power plant domains) and the authors give evidence that these methods of design appear to be both useful and needed in collaborative crew interface areas. However, they also point out a challenge to this community. They suggest that improvement in a design could also incur by using a 'conventional' human factors methodology and thus the overall worth and the specific components within their ecological interface design methods remain somewhat questionable. They suggest that the value of application lies in detecting where these methods best contribute to the *human engineering process*. They go on to point out contributions from this perspective and suggest a principled approach to design. Yet there still remains the question of "how efficient / how effective" the method is in translating information requirements into interface content, structure, and form. This will remain a daunting challenge to this community for years to come.

Other papers dealt with different views and concerns of methods and tools. The paper by McNeese & Rentsch (this volume) specifically looked at methods by which the collaborative interdependencies and social aspects of work could be elicited and assessed. Likewise, Perusich (this volume) presented an innovative method and representation formalism, *fuzzy cognitive maps*, for representing collaborative interdependencies. He makes a cogent argument for utilizing the fuzzy cognitive map as a diagnostic tool for understanding "friction" in team decision making. The paper evaluates this approach by applying it to a Scud-hunting scenario and shows it has value in identifying and discerning discrepancies in shared situational awareness. The paper by Sykora, Bahbouth, Radova, Dvorak, and Podivinsky provided stimulation and gives a good example of several integrated methods (e.g., the use of paper-pencil surveys with physiological analysis and video-audio recording) to model the aircrew under stress. As in the Perusich modeling representation their techniques also incorporate the use of fuzzy set concepts and provides a unique treatment of dynamic *socimetric methods* for use in studying intra-and-inter group relationships. Other papers (Oser, Dwyer, Cannon-Bowers, and Salas, this volume) looked at training frameworks and methods (event-based scenario training). In addition to these examples, the following methods and tools can be found throughout this volume; process outcome mappings, boundary constraint analysis, invasive data capture, event sequences, a/v recordings, protocol analysis, questionnaires and survey instruments, GOMS, ESDA, MacSHAPA, collaborative task analysis, behavioral checklists, participatory design, ethnographics, ethnomethodologies, prototyping, disfunction analysis, group concept mapping, function analysis, interaction process analysis, sociomaps, computer-augmented communication metrics, cognitive task analysis, and observation techniques. This is not a complete listing but it shows the breadth of methods presented at the symposium. One of the issues that scientists and practitioners are faced with is the question of the proper technique, method, or tool to use given respective strengths and weaknesses according to the situation. Answers to these questions require an intimate familiarity with a variety of tools and knowing when to use the right tool for the right job. There is much concern about potential misuse of methods and tools.

PSYCHOLOGICAL MEASURES / ASSESSMENT

This element is approached by varying levels of conceptualization and tends to skew more towards psychometric considerations, for this symposium. However, the paper given by Hollnagel (this volume) is a good treatise on understanding the intrinsic and mutual relationship between models and measurement. At the heart of his argument is that a proposed measure must capture the 'orderliness of performance' of a joint system. He goes on to look at three concepts: that measures must be possible, reliable, and valid, but indicates that measures rarely meet these requirements. He also analyzes the differences in measures that are derived from articulated

(theory-driven) versus folk models of everyday phenomena (theory-begging). Folk models are considered incomplete, focus on description rather than explanation, and are difficult to prove wrong. Hollnagel incisively distinguishes how measures differ as they derive from these alternative models and how this is useful for collaborative crew performance research. Specifically he relates these ideas to the elements of control and performance and presents the Contextual Control Model (COCOM) as a way of describing and predicting human behavior. In this sense COCOM is a framework for assessment and measurement that can be used for a variety of application domains. One of the challenges that Hollnagel points out is that of interpreting measurements after they have been taken. Perhaps this is especially true when considering psychometric measurement in team settings.

A majority of the papers in this element dealt with issues of measuring individual differences (Callister, Percival, Retzlaff, & King, this volume); personality characteristics (King, Callister, and Retzlaff, this volume); compatibility factors (Kay and Dolgin, this volume), and cognitive ability (Hoffmann & Kay, this volume); as related to collaborative crew domains. Some of the specific crew measures / assessments mentioned are: decision making errors, frequency of critical events, vital signs of patients, SWORD, ZEIS, RMPA, TLX, MC ISA, physiological parameters (e.g., heartbeat), SAGAT, cumulative fatigue, TMSS agreement and accuracy, event-driven measures, FIRO-B, COGSCREEN, and ALAPS to name a few. Some of the issues that need considered are the extent to which gender and cultural orientations will influence multinational collaborative efforts (King, Callister, and Retzlaff). As a crew is formulated to perform a mission, many of the individual and personality factors become very active under conditions of stress or extreme environments (see McNeese and Rentsch as an example of space operations on-board the Mir). The paper by Hoffman and Kay shows that in crew resource management, cognitive ability remains a strong predictor in shaping individual performance into effective crew performance. One of the conclusions in this area provided by Kay and Dolgin is that advances in the selection of individuals must be matched with assessments in selecting and forming effective teams. The challenge is determining how unique individuals with specific capabilities for performance can be fused together to increase collaborative crew performance through the use of valid assessments and measurements.

FIELDS OF PRACTICE

Most of the symposium papers were relevant to complex operational environments. In fact, an accumulated listing shows the following contexts in evidence; to name a few: NIMROD MR-4, air warfare centers, CC-130 Hercules aircraft, reconnaissance land vehicle, trauma resuscitation, TORNADO aircraft, international space station, AWACS command and control, unmanned air vehicles, submarine attack centers, air tasking orders, Mir space station, in-flight air emergency in commercial air liner, air defense crews, air traffic controllers, rotorcraft pilot's associate, suppression of enemy air defense mission, composite air operations, and a DC-9 crash. Some of the papers focused on the field of practice as the primary source of data while others identified the field / context but went on to some other aspect of research which did not specifically take an ecological or ethnographic perspective. Certainly, the way one melds the field of practice with the practitioner has overwhelming implications for what transpires in collaborative crew performance. This relates to many of the issues / challenges brought up in this summary (as to where a researcher begins and ends their activities) as well as the consequences of these activities upon design artifacts and advanced support systems.

There are two exemplary papers that demonstrate the value of the field of practice as a primary source of data. The first is the keynote presentation by Bannon (this volume) which looks at some of the philosophies and implications of embracing *social ergonomics* in the design of human-computer interfaces (HCI) and CSCW systems. The second paper by Xiao (this volume) provides a solid look at how to address complex medical domains through the emphasis of understanding workers by looking at their context of work. The paper by Xiao also describes useful methods of how to go about this form of ethnographic study.

The Bannon paper, perhaps more than any other in this volume, is important to read and reflect on as it looks at the history of a variety of influences on how researchers have attempted to study, design, and impact technological systems with the human in mind. This is specifically developed for computer systems and cooperative work. His thesis documents how traditional views (e.g., human factors and cognitivist approaches) may not be sufficient to adequately support studies and designs that capture the *social ergonomics* of work. He refers to the term, *work-context gaps*, to denote the type of elements that are missing when HCI-cognitivist communities employ their typical methods. These gaps are representative of what can happen when only *the practitioner* is highlighted but the field of practice is diminished. The field of practice (i.e., the complex operational environment which weaves the fabric of this symposium) provides many of the socio-cultural signatures that constrain work, workers, and the objects of work. Bannon proposes that much of the research in the CSCW community attempts to focus on supporting work environments via the use of ethnomethodologies. He uses the example of the air traffic control domain to show the ramifications of his observations. One of the very key points that Bannon makes is that different communities that essentially study similar phenomena must begin to make contact and become aware of their respective competencies. He points to the value of cognitive systems engineering (although not particularly used in the CSCW area) as well as the Francophone researchers in Europe (de Montmollin, Thereau, and de Keyser) who distinguish between actual work practice and normative accounts of work.

The paper by Xiao examines a specific field of practice in the medical domain, trauma patient resuscitation, in terms of human activities that are captured *in situ*. The paper utilizes methods involving video analysis such as expert commentary, coding of verbal communications, and event

flow analysis. The paper shows the importance of collaborative activities in real world settings while highlighting the temporal scale of work in this domain. A valued conclusion that Xiao makes is that basic collaboration *in situ* necessarily requires *tool using* and that future studies of complex systems must attune framework, methods, measures of assessment, and descriptive languages to mesh with the observation of actual work settings.

Other papers of special note in the symposium that specifically highlight unique field of practice insights with a broad research base are Elliott, Schiflett, and Dalrymple; Kirschenbaum; Campione, Brander, and Koritsas.

ADVANCED TECHNOLOGIES

The final element in the organizing structure provides a intriguing look at different kinds of technologies, the integration of multiple technologies around crew-centered concerns, and the design of specific collaborative technologies as a function of understanding the collaborative constraints and capabilities of crew members. Some of the different technologies mentioned across papers are team information displays, decision support systems, team training aids, 3 dimensional sound, CSCW technology, multiple information spaces, collaborative support systems, common interaction environments, software mediators, avatars, mobile computing, and adaptive aiding.

Advancing technologies through an understanding of the collaborative crew (for a given domain) is one of the main challenges presented at this symposium. Several researchers referred to the issues and challenges of developing designs for collaborative environments that would be adaptive to the emerging complexities that a crew must address. There are two papers in particular that emphasize some of the advances in collaborative systems and adaptive aids that I would like to point the reader to. The Duffy paper (this volume) is an excellent example of how collaborative technologies have been integrated to impact new avenues of how crews work together, It provides a good synthesis of MUD / MOOs, groupware, electronic whiteboards, and collaborative infrastructure; all in support of high-level planning and decision making activities in command and control arenas. It is most useful as it shows the integrative strand of coupling these technologies together to support, enhance, and adapt collaborative activities. The other paper by Taylor (this volume) is very good in reviewing much of the history and work in the human-electronic crew and describes many of the problems and challenges which need considered in crew-centered design of associated systems. It is progressive in that it begins to look at the electronic associate as a collaborative crew member. The implications of this conceptualization are drawn out through several dimensions and are valuable for any researcher who plans to develop technology that adapts with the crew and the environment. The amount of papers that are indexed to advanced technological systems are vast. One key issue is how collaborative infrastructures (as manifest in the new computer architectures) meet crew needs and concerns. The Thody and Ross paper (this volume) looks at some of these questions as they evaluate unique collaborative crew performance requirements in the design of a future land reconnaissance vehicle.

CONCLUDING REMARKS

As one can clearly see from an elaboration of the six preceding elements, there is much to glean from this symposium. A robust and broad sampling across many areas of interest and concern has emerged. It is exciting to have a symposium that illuminates so many innovative ways of conducting research and opportunities for approaching problems within crew collaboration. In contrast, one can also concomitantly be overwhelmed by conflicting views, alternative methods of assessing the same problem, and just the pure volume of knowing where (and how) to begin and end a research / development project. I think the key point noted by Dr. Bannon is worth repeating. He suggested that there is value in trying to complement "your approach" with what other communities are doing as they evolve the intersection of: a) the design of new collaborative systems, b) human-oriented approaches to system design, and c) work-centered practices.

Yet, I submit that one of the main problems we face as scientists is that of the "difficulty of convergence". At the core of *who we are* and *what we do* is the notion that we are first and foremost individual scientists. With this comes inevitable levels of variance associated with differing motivations, values, priorities, judgments, biases, and basic differences. We are shaped at the next layer out by the mores of the research groups we participate in and acknowledge. The individual and research group layer is further shaped by the organizational constraints that each of us work in. When we multiply the shaping of these layers together it is often the case that in fact there are many irreconcilable boundary constraints in working multiple perspectives together. Irreconcilable differences may show up as imbalance, high levels of variability, questions on where to begin projects, and not knowing what is doable (inability to make tradeoffs). When we put our thoughts together at a symposium such as this one, we typically find that our papers (or viewpoints) can be conflictual, competitive, complementary, or cooperative with what others present or with other communities of practice. I have made the analogy that we are like missionaries with apprenticeships, adapting through and with a community of practice. When we ask about how we should fit Collaborative Crew Performance in Complex Operational Systems together, it is similar to interlocking parts of a puzzle with only partial vision of the pattern to guide construction. Where we fit and where we are headed is a function of *mutual learning*, the willingness to keep our senses focused on many options, and the ability to openly share the social construction of a joint vision.

As we continue our quest for knowledge in these areas, it is important to maintain awareness with respect to the following queries to enable a cogent research path and agenda:

• What have we learned? (applications and research)

• What are we trying to learn more about? (key research variables)

• How are we going about learning more? (frameworks, methods, approaches)

• How are we using / implementing what we learn? (feedforward and design)

• How are we using what we learn to leverage new research? (feedback / scientific inquiry).

Answers to such queries will set the stage for trends and agendas in collaborative crew performance in complex operational environments as we approach the new millennium.

REFERENCES

1. Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*. New York: North Holland.

2. Rasmussen, J., Pejtersen, A, M., and Goodstein, L. P. (1994). *Cognitive systems engineering*. New York: John Wiley & Sons.

KEYNOTE ADDRESS

NIMROD MRA4 – COLLABORATIVE CREW PERFORMANCE CHALLENGES

by

Wing Commander N J Davis BSc MRAeS MRIN RAF
OR 22(Air)
Whitehall
London SW1A 2HB
United Kingdom

I would like to thank the Co-Chairpersons, Dr Ken Boff and Jo Davies, for the opportunity to give this keynote address to the RTA Human Factors and Medicine Panel's first symposium on Collaborative Crew Performance in Complex Operational Systems and to start off your weeks deliberations. May I say that it gives me an enormous satisfaction to play a small part in an organization, which provides a unique structure for international cooperation in aerospace research and development, particularly human factors and medicine.

What I would like to try and do in the time that I have available is to give you an operational perspective of the challenges we are likely to face as we introduce the new Nimrod Maritime Reconnaissance and Attack Aircraft, designated Nimrod MRA4, into operational service with the Royal Air Force. But before I do that, I am quite sure that many of you here in the audience have neither heard of, nor even seen a Nimrod Maritime Patrol Aircraft. Therefore, I will present you with a short history of the aircraft, briefly outline its operational roles, describe the new Nimrod MRA4 and finally address the crew performance challenges we expect to face.

In July 1964, Air Staff Target Number 381, which defined the requirements for a maritime reconnaissance aircraft to replace the Shackleton by 1968, was issued for the design, research, development and construction of two prototypes and 38 production aircraft. The price was 100 million pounds sterling and the aircraft, by now named Nimrod after a character in the biblical Book of Genesis described as "a mighty hunter in the eyes of God", was planned to remain in service until 1995. The first production aircraft flew on 28th June 1968 and, whilst for the pilots and flight engineers it was all a wonderful new toy, for the tactical team a lot of equipment was carried forward from the Shackleton. It was not until 1979, with the introduction of the Nimrod MR2, that the tactical team saw major equipment improvements in the radar, acoustic and Electronic Support Measures sensors. The new equipment was installed in the expectation that the aircraft would be replaced during the mid-1990s, but it soon became apparent that funding for a replacement aircraft would not be available, so plans for a mid-life update were formulated. Towards the end of the 1980s, with changing Royal Air Force priorities, and the coincidental development of the Lockheed P-7 in the United States, it became possible for the Nimrod to be replaced in the time-scales originally envisaged. Work on the mid-life update therefore ceased. Unfortunately, soon after this decision was taken, the Lockheed P-7 project was cancelled. Work on a new Staff Requirement now began in earnest.

Air Staff Requirement Number 420 was issued, initially calling for 25 Replacement Maritime Patrol Aircraft with an In Service Date of 2003. As well as aircraft, the Requirement included the supply of simulators, training packages and infrastructure enhancements. Although many companies expressed initial interest, only four serious competitors came forward. The European contenders were Dassault Aviation with the Atlantique 3, an improved version of the Atlantique 2 currently flying with the French Navy; and British Aerospace with Nimrod 2000, a significantly rebuilt and reequipped Nimrod MR2.

From the United States Loral proposed the Valkyrie, using new avionic systems in refurbished P-3A airframes that would be recovered from desert storage; and Lockheed projected the Orion 2000, a much modified, new-build P-3 airframe, that would be fitted out in the United Kingdom. At last, with the announcement on 27th July 1996 of the decision to award the contract to British Aerospace, the Royal Air Force finally had a clear way ahead. The contract, now for only 21 aircraft and worth just over two billion pounds sterling, was eventually signed with British Aerospace on 2nd December 1996.

There is no doubt that the threat posed by the former Soviet and Warsaw Pact Navies has greatly diminished. Todays threat does not come from one large overpowering country, rather from many smaller countries, for example, third world countries with the political and military will to try to enlarge their territories rapidly. The threat could come from a combination of surface warships, conventional and nuclear submarines. Today, Maritime Patrol Aircraft are tasked with monitoring these important naval assets in much the same way they used to keep watch on the former Soviet Fleet.

The Maritime Patrol Aircraft, regarded more now as a multi-role surveillance platform, operates primarily in three key roles: Anti-Submarine Warfare; Anti-Surface Unit Warfare; and Search and Rescue.

Anti-Submarine Warfare is a role that relates to the prosecution of submarine contacts. The submarines operating environment and modes mean that submarine detections can be made either above or below the surface of the sea. This, in turn, drives the requirement for an anti-submarine platform to have a sophisticated range of sensors and a reliable, advanced means of coordinating the sensor outputs. Once a submarine has been detected by the platform, the ultimate aim of the crew is to then localize and track the target, or even destroy it.

Anti-Surface Unit Warfare is a role that relates to the prosecution of surface shipping, both military and commercial. The military surface ships operating environment means it is relatively vulnerable to aircraft, and it therefore compensates by maximizing the threat it can pose to aircraft. Such units can carry a wide range of technologically advanced, long-range anti-aircraft weapon systems. Non-combatant vessels must rely on the remoteness bestowed by millions of square miles of ocean, as their only defence against intervention by aircraft. The combination of the need to search large areas of ocean, and the requirement to stand-off from potentially hostile contacts, drives the requirement for an anti-surface warfare platform to have one or more sophisticated, long-range search and classification sensors, such as radar and Electronic Support Measures equipment. Once a surface contact has been detected by the platform, the aim is to shadow the target, or even destroy it.

Let me now turn to the final role of Search and Rescue. Search and Rescue is an activity that takes place during peace and war. It traditionally takes the highest priority in any list of tasks, and it varies in nature from incident to incident. A Nimrod is always held at one hours and six hours readiness, capable of dropping Search and Rescue equipment or acting as a Scene of Search Coordinator for incidents such as oil rig disasters. It is as a command and control coordinator that the Nimrod excels in the happily, rare cases of a major disaster at sea. The Piper Alpha oil platform tragedy is a well remembered instance of rescue helicopters and ships converging from all points of the compass into a confused situation. Although such events are technically managed by a Rescue Coordination Centre, the responsibility is preferably delegated to someone at the scene. Nothing is better equipped than the Nimrod functioning as a director of operations and as an airborne relay station.

The Nimrod MRA4 is the platform that has been chosen to fulfil the three key roles of Anti-Submarine Warfare, Anti-Surface Unit Warfare and Search and Rescue for the Royal Air Force. Although based on the existing Nimrod MR2, over 60% of the airframe will be of new construction. The aircraft will have a new wing, a new and wider undercarriage, and will be powered by new BMW/Rolls Royce engines, which will offer some 25% better fuel consumption and 30% more power than the current Nimrod engines. The cockpit has been designed for two-man operations with a high

degree of automation installed. The 'glass cockpit' will be similar to that found on the A320 and A340 variants of the European Airbus and will have full-colour, liquid-crystal display screens presenting flight system, navigation, fuel and engine data, as well as a wide range of tactical data. The aircraft will have its own set of integrated stairs and its own auxiliary power unit, capable of providing all the power requirements, prior to engine start, for tactical and sensor equipment checks, flying controls and hydraulic system checks, and conditioning of both cabin and equipment. The radar will be the Racal Searchwater 2000, a development of the current radar on the Nimrod MR2. The Electro-Optical Surveillance and Detection System will be based on the Northrop Grumman Night Hunter system and will provide an infrared and low-light television capability. The Electronic Support Measures equipment will be the Elta EL 8300(UK) and the acoustics system, designed for two operators, will be a development of the Computing Devices of Canada AN/UYS 503 processor, capable of monitoring up to 64 sonobuoys. The communications suite will be capable of handling V/UHF and HF radios, Links 11 and 16, and Satcom. Bringing all the tactical data together, to enable the crew to efficiently operate the aircraft as a weapon platform, is the Tactical Command System, which is being developed by Boeing. The aircraft will also be fitted with an extensive defensive-aids system developed by Lockheed Martin. In addition to the current mix of sonobuoys and survival equipment, the Nimrod MRA4 will be capable of carrying a wide range of torpedoes, air to surface and air to air missiles. Finally, only two sensors currently installed in the Nimrod MR2 will be carried forward onto the new aircraft: the Mark One Eyeball and the Magnetic Anomaly Detector. To manage this vast array of highly sophisticated equipment is a team of just ten people, three less than on the current Nimrod.

British Aerospace intend to use the first three Nimrod MR2 airframes to undergo conversion to the MRA4 standard as system development or trials aircraft. The first flight is currently scheduled for September 2000 and the first flight with a fully integrated mission system is scheduled for March 2001. The first operational Nimrod MRA4 should be delivered to Royal Air Force Kinloss in July 2002, while the In Service Date, defined as the date when seven aircraft are fully operational, is currently scheduled for April 2003. The twenty-first and final aircraft should be delivered to the Royal Air Force in 2006. It is worthy of note that it will be just over four years from contract award to the first flight, a truly remarkable achievement when you compare it to other high technology aircraft development programmes.

Before I finally address the challenges to be faced by the crew of a Nimrod MRA4, I think it is important, particularly amongst such a distinguished audience, to briefly mention human factors and the importance it has in this project. As you are all aware, human factors is the study of man in his working environment. The human in a system is often the limiting factor in its overall performance: that is, the system will only perform as well as the operator. In the military, personnel are expected to operate in harsh conditions, under high stress with sophisticated technology. Designing the equipment around the operator to optimize the overall performance of the system would therefore be a logical move to make. However, experience has shown that human factors issues are often not addressed early enough in the procurement process with the consequence that pre-production prototypes often have significant problems. Modifying designs at such a late stage in the equipment development process is often very expensive and usually causes a delay to the project. Now as far as the Nimrod MRA4 project is concerned, human factors issues have been addressed early in the procurement process, and will continue to be addressed throughout the life of the project.

There are perhaps only two areas that will present a real challenge to the crew of Nimrod MRA4: the two-man cockpit and the overall crew workload. Since the flight engineer has effectively been replaced by automatic management systems, and as the pilots are now going to be responsible for the aircraft navigation, the cockpit is likely to be a very busy area. Moreover, with the vast amount of data that will be gathered from all the aircraft's sensors, crew communication, particularly intercom discipline, is going to be crucial if the total weapon system is to function efficiently and effectively. Consequently, workload issues, particularly in a

multi-crew environment, will be significant. Mr Alan Felstead from British Aerospace will be addressing Nimrod workload issues during this afternoon's session.

Although described as a Maritime Patrol Aircraft, the Nimrod MRA4 will be a true multi-role combat aircraft. With only 21 aircraft and a crew of ten, the force will be slimmer, but I believe a more capable and potent one for the Royal Air Force to take into the next millennium. The enemy should take heed, for in the eyes of God, the Mighty Hunter reigns supreme.

Finally, it is clear from the conference agenda that the papers to be presented over the next three days will certainly explore the methods, techniques and tools associated with the design, development and evaluation of multi-crew systems. I look forward to a most interesting symposium and would now like to hand over to the first session chairmen. Thank you.

TACTICAL DECISION MAKING UNDER STRESS (TADMUS): MAPPING A PROGRAM OF RESEARCH TO A REAL WORLD INCIDENT-THE USS VINCENNES.

Joan Hall Johnston, Janis A. Cannon-Bowers, & Eduardo Salas
Naval Air Warfare Center Training Systems Division Code 4961
12350 Research Parkway
Orlando, FL 32826-3275, USA

ABSTRACT

Such events as the one involving the USS Vincennes, where the decision to initiate countermeasures was the incorrect one, have focused attention on the human factor in decision making. The objective of the TADMUS program has been to apply developments in decision theory, individual and team training, and information display to the problem of enhancing tactical decision quality under conditions of stress. Sponsored by the Office of Naval Research, TADMUS is in its 8th year as a cooperative program in human factors and training involving SPAWAR Systems Center, San Diego, NAWCTSD, as well as other Navy, industrial, and academic organizations. The technology is being demonstrated and evaluated in the context of surface ship air warfare scenarios. This address will describe how the TADMUS program was founded and how it has progressed on a variety of R&D issues having to do with advanced training and human factors in order to address real world problems.

INTRODUCTION

On July 3, 1988 an AEGIS cruiser named the USS VINCENNES accidentally shot down a commercial aircraft, Flight 655, over the Arabian Gulf killing 290 people. A number of consequences resulted from this catastrophe that included the initiation in 1990 of an Office of Naval Research sponsored research and development program named Tactical Decision Making Under Stress (TADMUS). The major program goal has been to prescribe empirically-based principles and guidelines to enhance team tactical decision making performance in complex, knowledge rich environments. Therefore, program objectives have focused on applying developments in decision theory, individual and team training, and information display to the problem of enhancing tactical decision quality under conditions of stress (Ref 1).

TADMUS is in its eighth year as a cooperative program in human factors and training involving the Naval Air Warfare Center Training Systems Division, Space and Naval Warfare Systems Center, San Diego as well as other Navy, industrial, and academic organizations. To date, the written products generated from this program in its first six years number more than 200 publications in the form of journal articles, technical reports, book chapters, and symposium proceedings. In addition, over 100 product transitions have been provided to the fleet training communities in the form of lectures, workshops, demonstrations and implementations of training tools, methods, and strategies. Finally, three large-scale advanced embedded training research initiatives have resulted (Ref 2; Ref 13). This paper describes how the strategic planning for the TADMUS program enabled us to design a roadmap for achieving a program of research accomplishments for advancing current and future developments of Navy combat team training.

BACKGROUND AND APPROACH

Three TADMUS program thrusts were identified for the purpose of meeting its program goal: definition and measurement of critical decision tasks; examination of stress effects on decision making; and development of training and simulation principles (Ref 1). Program success would be assessed by the emphasis on empirical research, rapid transition of research products to the Navy training community; and research findings that would support the development of advanced team training technologies. Below is a description of program requirements that defined our strategy for meeting program goals.

Technical Advisory Board. A Technical Advisory Board populated by senior researchers and high level fleet representatives was appointed to review technical progress twice per year until program completion. Without a doubt, the frequent meetings ensured that researchers remained focused on areas of work that were going to bear fruit; overall, the Technical Advisory Board helped keep the research on track.

Subject Matter Expertise. Inclusion of Navy subject matter expertise, training commands, and ship teams in the design and development of research experiments and products was an integral part of the TADMUS plan. To date, virtually hundreds of active duty officers and enlisted men and women have played a role in shaping the research products so they would have a realistic and substantive impact on current and future Navy training.

Defining and Designing a Realistic Team Task. The VINCENNES incident was due, in part, to errors resulting from the "Air Warfare" (AW) team interactions, therefore, the research domain was to examine the AW task in a ship's Combat Information Center (CIC). The CIC is the central information processing and tactical decision making area for a

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

surface combatant (Ref 4). The main focus of research would be the Air Warfare team on an Aegis capable ship, which is composed of the Commanding Officer, Tactical Action Officer, Air Warfare Coordinator, Tactical Information Coordinator, Identification Supervisor, and the Electronics Warfare Supervisor. During air warfare, the AW team performs a series of tasks including detecting, tracking, identifying radar contacts, taking action on these contacts, and performing battle damage assessment (Ref 4).

Once the operational task was determined, then choices had to be made regarding a team-based research testbed. High costs and lack of access meant that utilizing a ship's CIC or a shore-based high-fidelity team trainer was out of the question. Therefore, a five-person networked PC-based system called the Decision Making Evaluation Facility for Tactical Teams (DEFTT) was developed to support research experiments. Although, as a result of a research tradeoff, DEFTT was low in functional fidelity, it was determined that the system presented the tactical problem with enough cognitive fidelity that we could "simulate" the actual AW team activities taking place in the real combat system (Ref 5; Ref 6).

Teams. Without a doubt, conducting team experiments was the most challenging of all the research tasks. We determined that most of the experiments would include five-person teams. Consequently, as with developing DEFTT, it was not realistic to expect our research participants to be highly experienced operators that had worked together as a team for an extended period of time (i.e., intact). Therefore, we developed a plan so that we could eventually gain access to such teams. We organized our research tasks to be conducted at three shorebased training commands where we could have access to novice, experienced, and very experienced trainees. In addition, we spent the first two years of the program at a high fidelity combat team training facility where we conducted numerous interviews and developed a database from which our research scenarios were developed. We were able to establish DEFTT at one of the school commands, and another command had already adopted it as a trainer for division officers, department heads, and prospective commanding officers. In addition, we established a low fidelity 3-person team trainer at a school command that allowed novice Navy recruits to participate in some of the basic research experiments. To date, over 280 teams of Navy trainees have participated in the research. Once we had established a reliable research protocol we were able to have intact and ad hoc experienced ship teams participate in the research (over 10 teams to date), as well.

Event-Based Scenarios, Stressors, and Measurement. To ensure that the innovative training could be evaluated, we developed a strategy—the Event-Based Approach to Training, or EBAT—to support research scenario design, measurement tool development, and operational stressor implementation (Ref 7; Ref 8). We structured two pairs of AW Arabian Gulf scenarios. Each pair was composed of one low and one high stress scenario, but both shared the same events. Stressors were defined as workload (e.g., added aircraft and ships) and information ambiguity (e.g., increased number of difficult problems to solve). Each scenario was developed with three significant "events" whereby individual and team behaviors could be specified by subject matter experts, and observed by trained raters (e.g., two hostile aircraft popup at close range to the team's own ship). The EBAT strategy thus provided a way to ensure we could assess individual and team performance with measurement tools that were designed to capture performance processes and outcomes (Ref 7). Next is a brief description of the tools as they related to EBAT.

The Behavior Observation Booklet was designed for assessing individual task processes. At each scenario event, we identified task requirements at the individual operator level so that they had a performance score for each event. An outcome score measured by the Sequenced Actions and Latencies Index represents the ability to perform the tasks correctly and on time. The Air Warfare Team Observation Measure assesses team level performance for information exchange, initiative, supporting behaviors, and communication. The team outcome measurement tool (Air Warfare Team Performance Index) assesses timeliness and accuracy as a team on the detect-to-engage sequence. As a result of developing these tools, we were able to assess the stressfulness of research scenarios, and to guide assessment of the impact of the TADMUS training (Ref 9).

The Training Research Agenda. In the beginning of TADMUS, syntheses of the research literature on decision making, teams, and stress were conducted to identify and develop innovative training strategies (Ref 1). Consequently, a research agenda was designed so as to test the individual effects of such training on enhancing skills in critical thinking for decision making, teamwork and team self-correction, handling stress exposure, and leadership (Ref 10). In addition, training strategies and methods were tested to assess the impact of part task training, cross-training, and instructor training to enhance performance feedback strategies (Ref 11; Ref 6; Ref 12). The data collection effort started in 1992 and continues through 1999 in order to establish the impact of an integrated training program with the decision support system developed by SPAWAR in San Diego. To date, empirical data has been collected from over 95 five-person teams (including experienced intact ship teams) and findings have shown training imposed significant improvements in performance, often at levels of 30-40 percent (Ref 10).

Product Transitions: Short-term and Long Term. Although an applied research program, emphasis was placed on ensuring the fleet would gain short and long term benefits from our

findings. In the short-term, numerous such activities as workshops and demonstrations have and continue to take place. As an example of building a strategy to transition training to the shipboard environment, we demonstrated empirical support for training teamwork skills (Team Dimensional Training) in the laboratory and then in the shipboard environment, whereby we eventually gained the support and endorsement of the afloat training group for incorporating and implementing Team Dimensional Training (Ref 12; Ref 13). For the long-term, we have initiated several advanced research programs to ensure that shipboard embedded training that includes our TADMUS training methods, tools, and strategies will be incorporated into new ship platforms of the 21st Century (Ref 2).

SUMMARY AND CONCLUSIONS

We have described the strategy we used to guide us on the roadmap for achieving the TADMUS objectives. The Technical Advisory Board, subject matter expertise, research task design and team participants, EBAT, the research agenda, and short and long-term transitions were crucial to ensuring the program's ongoing success. In conclusion, the vital components in all of these tasks were: (1) ensuring that empirical results were based on a reasonably realistic task that included team participation, and (2) that fleet participation—the customer—had input throughout the program.

REFERENCES

1. Cannon-Bowers, J.A., Salas, E., & Grossman, J.D., "Improving Tactical Decision Making Under Stress: Research Directions and Applied Implications", in the "International Applied Military Psychology Symposium", 1991.

2. Naval Air Warfare Center Training Systems Division. "Research and Development Project Summaries". NAWCTSD, October 1997.

3. Zachary, W., Bilazarian, P., Burns, J., & Cannon-Bowers, J. A., "Advanced Embedded Training Concepts For Shipboard Systems", in the "Proceedings of the 19th annual Interservice/Industry Training, Simulation and Education Conference", December 1997, pp. 670-679.

4. "CG 52 Class Combat System Doctrine", Naval Surface Force. COMNAVSURFLANT Instruction C3516.39. March 1989.

5. Holl, R. E. & Cooke, J. R., "A Recognition-primed decision (RPD) model of rapid decision making", in the "Proceedings of the 11th annual Interservice/Industry Training Systems Conference", 1989, pp. 337-342.

6. Kirlik, A., Rothrock, L., Walker, N., & Fisk, A. D., "Simple Strategies or Simple Tasks? Dynamic Decision Making in "Complex" Worlds", in the "Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting", Human Factors and Ergonomics Society, 1996, *1*, pp. 184-188.

7. Hall, J. K., Dwyer, D. J., Cannon-Bowers, J. A., Salas, E., & Volpe, C. E., "Toward Assessing Team Tactical Decision Making Under Stress: The Development of a Methodology for Structuring Team Training Scenarios." in the "Proceedings of the 15th Annual Interservice/Industry Training Systems Conference", December 1993, pp. 87-98.

8. Oser, R. L., Cannon-Bowers, J. A., Dwyer, D. J., & Salas, E., "Establishing a Learning Environment for JSIMS: Challenges and Considerations", in the "Proceedings of the 19th annual Interservice/Industry Training, Simulation and Education Conference," December 1997, pp.144-153.

9. Johnston, J. A., Smith-Jentsch, K. A., & Cannon-Bowers, J. A., "Performance Measurement Tools for Enhancing Team Decision Making", in M. T. Brannick, E. Salas, & C. Prince (Eds.), "Team Performance Assessment and Measurement: Theory, Methods, and Applications", Mahwah, NJ: USA, LEA, 1997, pp. 311-327.

10. Salas, E., Cannon-Bowers, J. A., & Johnston, J. H., "How can you turn a team of experts into an expert team? Emerging training strategies", in C. Zsambok & G. Klein (Eds.), "Naturalistic decision making ", Hillsdale, NJ, USA, LEA, 1997, pp. 359-370.

11. Duncan, P. C., Rouse, W. B., Johnston, J. H., Cannon-Bowers, J. A., Salas, E., & Burns, J. J., "Training Teams Working in Complex Systems: A Mental Model-Based Approach." in W. B. Rouse (Ed.), "Human/technology interaction in complex systems", *8*, 1996, pp. 173-231.

12. Smith-Jentsch, K.A. Johnston, J.H., Cannon-Bowers, J.A., & Salas, E., "Team Dimensional Training: A Methodology for Enhanced Shipboard Training", in the "Proceedings of the 19th annual meeting of the Interservice/Industry Training, Simulation and Education Conference", December 1997.

13. Smith-Jentsh, K.A., Payne, S. C, & Johnston, J.H. "Guided Team Self-Correction: A Methodology for Enhancing Experiential Team Training", in K.A., Smith-Jentsch (Chair), "When, How, and Why Does Practice Make Perfect?", Paper presented at the 11th annual conference of the Society for Industrial and Organizational Psychology, 1996.

An Exploratory Application of Ecological Interface Design to Aircraft Systems

D. Beevis[1], **K. Vicente**[2] and **N. Dinadis**[2],
[1] Defence and Civil Institute of Environmental Medicine,
1133 Sheppard Ave West, P.O. Box 2000, Toronto, Ontario M3M 3B9, Canada

[2] Cognitive Engineering Laboratory, Department of Mechanical & Industrial Engineering, University of Toronto,
5 King's College Road, Toronto, Ontario M5S 3G8, Canada

1. SUMMARY

Ecological interface design (EID) is a theoretical framework for designing Operator:Machine Interfaces (OMIs) that tries to integrate different kinds of representations into a common interface based on two concepts from cognitive engineering: the abstraction hierarchy and; the skills, rules, and knowledge framework. The abstraction hierarchy is a multilevel knowledge structure that can be used to develop physical and functional models of systems as well as the mappings between them. The skills, rules, knowledge framework provides principles for information to support those three levels of behaviour. To date, most applications of EID have been to process control. In order to explore the applicability of EID to aircraft systems and to build on previous work, an exploratory application was made to the systems of the CC-130 Hercules aircraft which are controlled by the Flight Engineer (FE). The project included: in-flight familiarization; a protocol analysis of FE tasks; preparation of an abstraction hierarchy of the CC-130 systems; definition of the interface content and structure, and; representation of the information in visual form. The outcome was a rapid prototype of an 'EID interface' for the CC-130 engineering systems that was evaluated by a focus group of Canadian Forces Flight Engineers. The study concluded that: the principles of EID can be applied to aircraft systems; EID needs to be supplemented by more specific design principles, and; EID can be integrated with such principles. Operator response to the prototype showed that the design of the OMI for one operator needs to take into account the responsibilities and functions of other crew members.

2. PREFACE

AGARD symposia have often identified a concern for the effective inclusion of human factors concerns in aircraft design and development. In one of the more recent reviews (1) it is concluded that "operational problems in the cockpit can stem from the process adopted in its design and development. Hence the effectiveness of a cockpit and the pilot within it bear a distinct relationship to the efficacy of the process that derived it. This is even more true for complex and integrated glass cockpits than for more conventional predecessors."

Guidelines exist for the application of human engineering in aircraft design and development. The usual approach recommended for the human engineering process is to start from a mission analysis and perform a functional decomposition to the point where operator tasks can be identified and from that design or select the necessary displays, controls and workspace (2, 3, 4 and Figure 1). The overall process parallels other system engineering activities.

From the viewpoint of the human engineering practitioner, concerns lie not in the efficacy of the overall process but in the effectiveness of the steps within it. Although the human engineering literature contains many approaches to improving the design of the aircraft:crew interface (see Refs. 5 & 6 for example) there is a lack of effective techniques for translating a functional description of an operator's tasks into an effective OMI design. The work reported in this paper is one of several attempts by the Defence and Civil Institute of Environmental Medicine to develop improved techniques for applying human engineering to projects sponsored by the Canadian Department of National Defence.

3. ECOLOGICAL INTERFACE DESIGN

Ecological interface design (EID) is a theoretical framework for designing OMIs that tries to integrate different levels of information into a common interface (7). In the context of the overall approach to human engineering provided by Figure 1, EID deals with the analysis of operator and maintainer tasks and with system and sub-system design. EID complements the other human engineering techniques which are required to complete the analyses shown in Figure 1, such as timeline analysis and operator loading.

The approach takes its name from its focus on the *interaction* between the human organism and its environment. In this respect EID has much in common with systems theory and the 'systems approach' to human factors. Because of the focus on the interaction between operator and machine, EID concentrates on the analysis of the operator's task environment rather than on a normative task analysis. Activity and task sequences are ignored in favour of the identification of three classes of information:

- the functional problem space of the operator
- the generic tasks to be accomplished by the operator
- the set of strategies that operators use to carry out those tasks.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

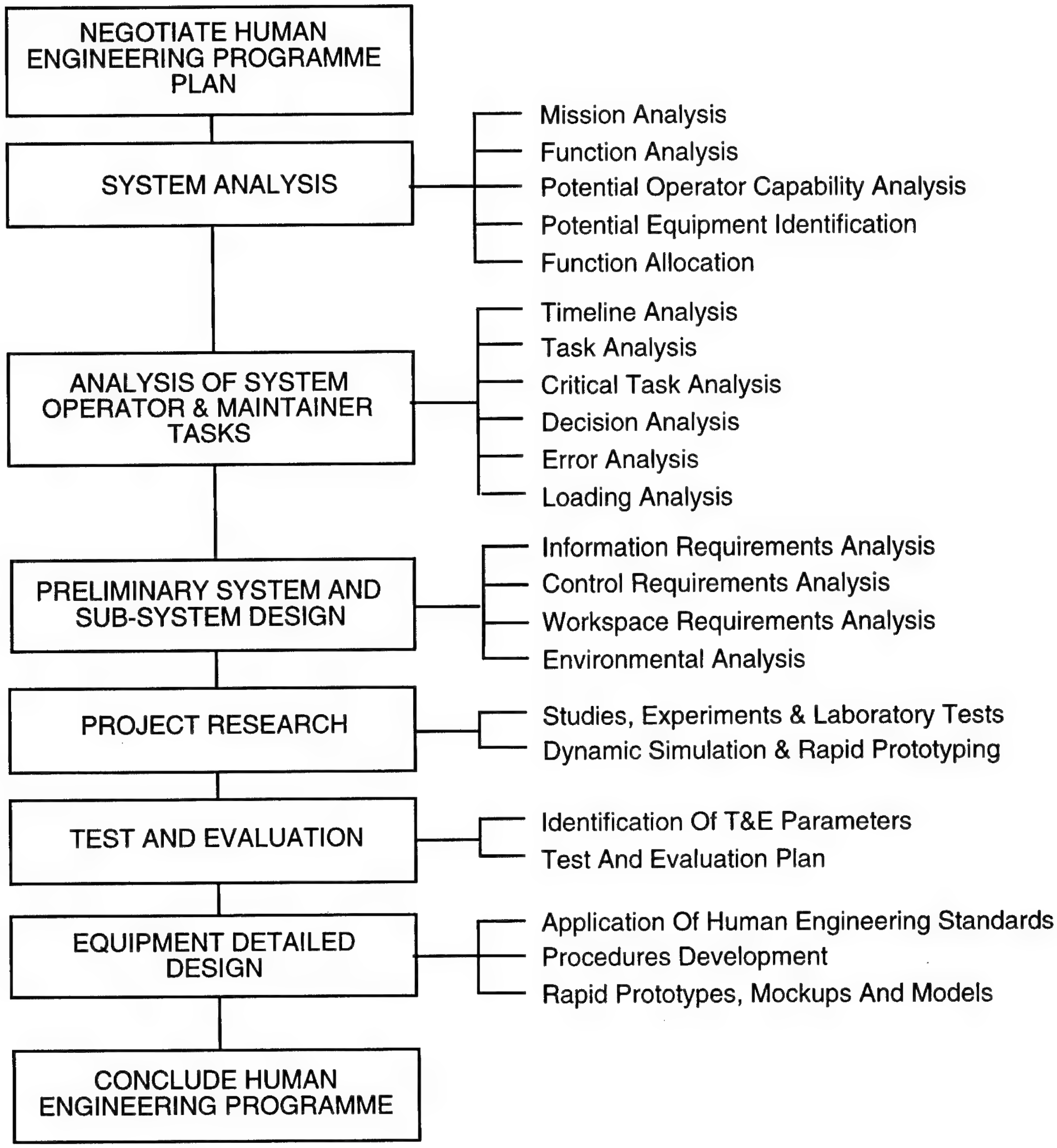


Figure 1: General human engineering programme model for advanced aircrew systems (STANAG 3994)

EID develops the information of the operator's problem space, generic tasks and strategies using two concepts from cognitive engineering research:

- the *abstraction hierarchy*
- the *skills, rules, knowledge framework.*

The *abstraction hierarchy* is a multilevel framework that relates the functional and physical aspects of a system. The abstraction hierarchy is similar to the decomposition of system functions and their translation into a 'means of implementation' that is part of the systems engineering process (4). The abstraction hierarchy can be used to map the relations between physical and functional system models and the whole system or parts of a system (the part-whole hierarchy) (Figure 2).

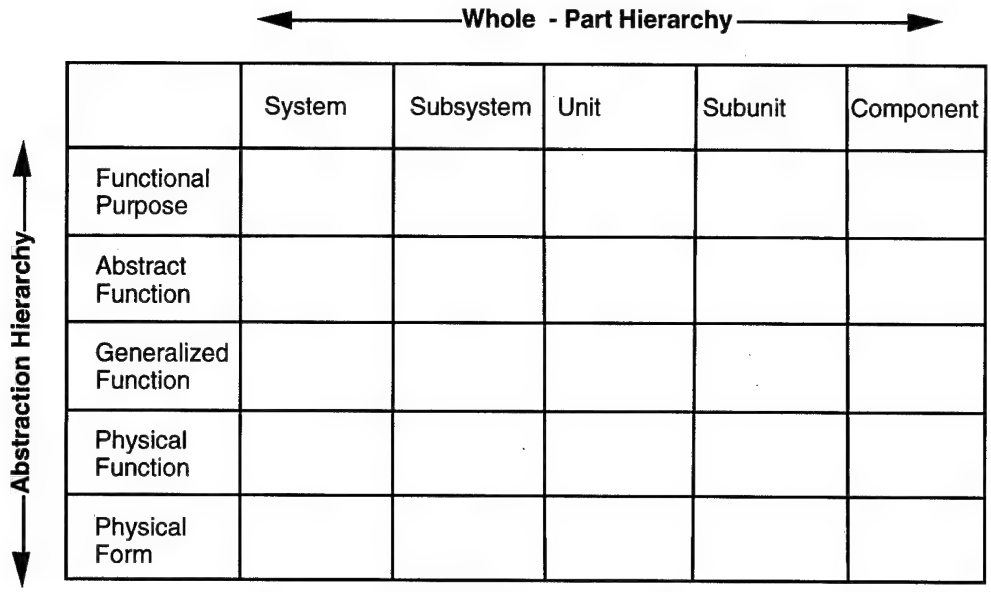


Figure 2: Generalized abstraction hierarchy for a system

Used in this way, the abstraction hierarchy can facilitate the understanding or diagnosis of system states (8, 9). It has been shown that operators can use information relating the abstraction hierarchy and the whole-part hierarchy for problem solving activities and that such information provides a basis for coping with events that are not only unfamiliar to the operator, but that may not have been anticipated by designers (8, 10).

The *skills, rules, knowledge (SRK) framework* (11) is based on the premise that humans work at three levels of behaviour (Figure 3):

1. skill-based behaviour involves sensory-motor performance which takes place without conscious control
2. rule-based behaviour is governed by 'stored' rules or procedures which have been learned or communicated
3. knowledge-based behaviour involves the formulation of goals and plans based on an analysis of the situation.

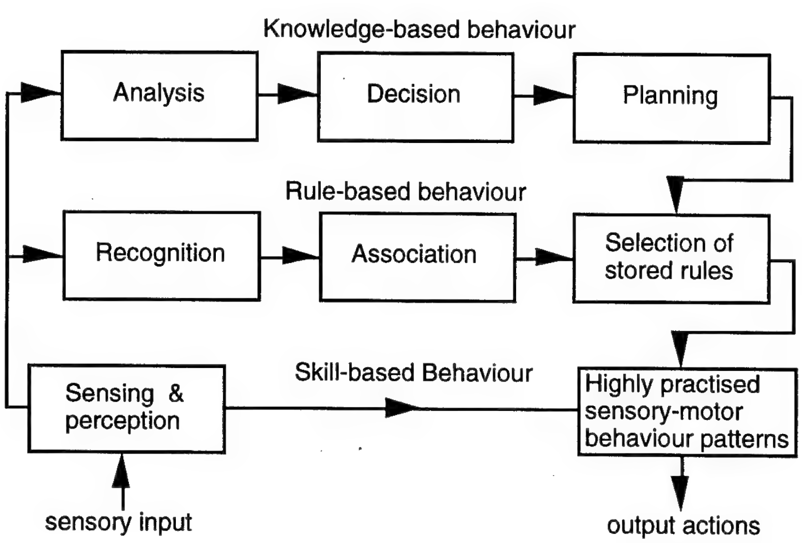


Figure 3: SRK levels of behaviour (after Rasmussen, 1983)

The SRK framework provides the following principles for interface design (7), all of which assume a computer-based OMI:

a) to support skill-based behaviour, operators should be able to act directly on the display, and the structure of the displayed information should permit 'chunking' of cues into higher level signals

b) to support rule-based behaviour, the interface should maintain a consistent one-to-one mapping between the perceptual cues provided in the interface and the underlying behaviour of the system (i.e., the display should show the relationships between system components that are operated at the skill-based level)

c) to support knowledge-based behaviour, (the formulation of goals and plans) the interface should provide information on the work domain in the form of an abstraction hierarchy (i.e., information on component, sub-system and system status and functioning and goals) which can serve as an externalized mental model to support problem solving.

4. APPLICATION OF EID TO AIRCRAFT SYSTEMS

Most applications of EID have been to process control, particularly power station control rooms. At the time the current project was initiated, EID had not been applied to aircraft systems although Vikmanis (12) had suggested using the SRK framework as a basis for analyzing pilots' tasks. In order to explore the feasibility and value of EID to aircraft systems design while building on previous applications, it was decided to apply it to the Flight-Engineer OMI of a multi-engined aircraft. The aircraft engineering systems (AES) of the CC-130 Hercules aircraft operated by the Canadian Armed Forces was selected for the study, after attempts to obtain access to, and design documentation on, more recent aircraft such as the latest generation of airliners were unsuccessful.

A literature review was first conducted on interface design methodologies and on current approaches to display design, including integrated displays, task-oriented displays, and emergent feature displays. It was concluded that most of the research had focused on issues associated with visual form and with providing information about the system morphology, rather than on the display of system functioning. Few guidelines were available for the latter area, where practice favours mimic diagrams that reflect the physical form of the sub-systems and units.

The next step was in-flight familiarization with CC-130 operations and Flight Engineer (FE) tasks. One of the authors (Dinadis[1]), flew some 90 hours of CC-130 operations. This, and reference to the CC-130 Crew Guide provided an understanding of the FE's roles, duties and responsibilities and of some of the more obvious limitations of the 1960's era displays and controls of the CC-130. The familiarization also provided the opportunity to observe some of the diagnostic decision making that FEs perform when some aircraft systems become unserviceable.

The flight observations were supplemented by a protocol analysis of FE tasks observed during two simulator training sessions in which aircraft faults had to be detected and remedied. The training scenarios involved simple faults which propagated through the aircraft systems and the observations identified the complexity of diagnostic procedures which involve understanding interactions between factors such as aircraft drag, fuel flow, and aircraft balance. The observations also indicated the potential value of parameter displays that would show trend or rate information.

The next step was the preparation of a matrix of an abstraction hierarchy/ whole-part hierarchy of the CC-130 Hercules systems, as shown in Figure 3. Not all cells of the matrix were useful for the analysis; those marked in grey were found not useful, those marked N/A were not applicable. The matrix provided the basis for identifying the information required for the various aircraft systems. For example:

- the functional purpose of the fuel system, was defined as being "to provide enough fuel to fly to a predetermined location"
- at the abstract function level this was expressed as mass flows of fuel in the system, subsystems (right and left wing fuels systems) and units (reservoirs, mass transports, energy sources etc.)
- at the generalized function level additional functions were included for the sub-systems and units, such as fuel filtering, fuel cavitation control, and engine turbine inlet temperature (TIT) scheduling
- at the same generalized function level, component information included 'fuel storage' 'flow maintenance' 'manifolds' 'drains; etc.
- at the physical function level the information included the arrangement and state of tanks, pumps, valves etc.

The above analysis drew heavily on prior experience in applying EID to power station control. Each item of information identified by the analysis was then categorized as:

- already provided
- can be measured with additional information
- requires manipulation of system configuration
- measurable on the ground only
- requires calculation
- not measurable
- automatic controller.

The analysis identified a number of important areas where information is not provided by the current OMI. In general, information is available at the physical function/ component level, but less and less information is provided for the higher functional levels.

[1] The majority of the CC-130 application study was conducted by Dinadis to fulfill the requirements of a Master's Degree thesis, under contract No. W7711-4-7234/001 with DCIEM.

	System	Subsystem	Unit	Subunit	Component
Functional Purpose					
Abstract Function					
Generalized Function					
Physical Function					
Physical Form			N/A	N/A	N/A

Figure 4: Abstraction/ Whole-Part Hierarchy developed for the project

The lack of higher level functional information requires increasing levels of effort from the FE to determine that information. For example, the only way to determine the flow in one fuel sub-system is to divert fuel through a meter by shutting off all crossfeed flows. Much of such information is derived currently by such 'work arounds' and calculations in the FE's logbook.

To develop an EID OMI, the definition of the interface content and structure provided by the analysis of the abstraction/ whole-part hierarchy matrix was translated into a visual representation of the state and functional information on the components, sub-systems and systems. An overall layout for the EID was developed which incorporated the various levels of information in the abstraction/ whole-part hierarchy (Figure 7).

The EID approach to displaying this information is to represent the underlying physics of a system. This activity again drew on previous experience in representing the goals and functioning of power generating stations and similar systems by displaying mass flows for fuel, oil etc., and energy balances for the engines (10). Mass flows were indicated using simple graphics that convey reservoir contents and flow rates; entropy state diagrams were used to display the status of each engine (Figure 5). The interpretation of the information requirements also drew on research conducted for aircraft systems displays, for example map representations of fuel-limited ranges (13) and polar star displays of system status (see, for example, Figures 6 & 8 and Ref. 14).

Despite having such 'models' or design metaphors available, the exercise made it clear that EID principles, in themselves, do not provide much detailed guidance on the implementation of the visual form of the necessary information in an OMI. Guidance for the design of visual form, in particular the separation of 'background' and 'foreground' layers of information was obtained from the work of Mumaw, Woods & Eastman at Westinghouse (15).

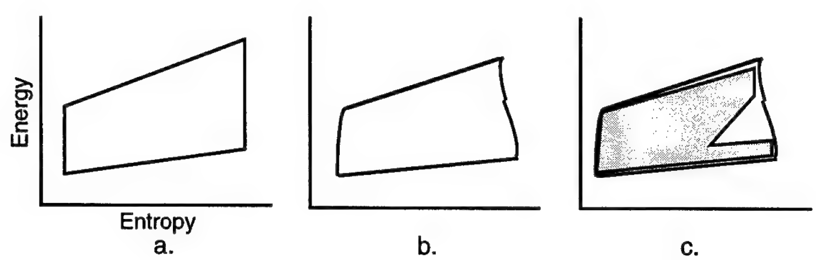
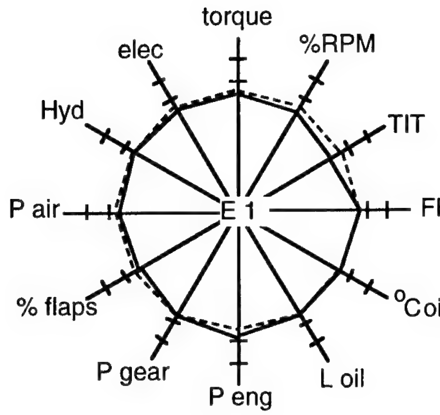


Figure 5: Entropy-state diagrams of:

a) a theoretical engine, b) normally operating engine, c) engine with low fuel flow



RPM = revolutions per minute
TIT = Turbine Inlet Temperature
FF = Fuel Flow
°C Oil = Oil temperature
L oil = oil level
P eng = engine oil pressure
P gear = gearbox oil pressure
% flaps = percent to which the engine cooling flaps are open
P air = bleed air pressure
Hyd = engine hydraulics status
elec = Engine electrical status

Figure 6: Polar star diagram of engine parameters

The EID interface for the CC-130 was developed using a VAPS® rapid prototyping system which allowed a limited representation of the dynamics of the various displays. A typical view of the final OMI is shown in Figure 7. Space precludes a detailed description of the interface, which is provided in the project report (16).

One barrier to the application of human engineering is that the necessary analyses are often considered to be labour intensive. Because of such concerns, the level of effort to apply EID in this project was logged. The amount of effort spent on the project (by an MSc-level student with no prior experience of aircraft systems) was as follows:

- Analysis of system 4 months
- Design of EID interface 1 month
- Programming VAPS 2 months

(The programming effort would probably have been much less for someone with experience in using VAPS).

5. EVALUATION OF THE EID OMI CONCEPT

At the outset of the project it had been intended that the final EID design would be evaluated by experienced FEs. Due to personnel reductions and a heavy workload among CC-130 operators it was not possible to use experienced personnel for man-in-the-loop studies of the VAPS prototype. Therefore, the EID prototype was presented to a focus group of seven Canadian Forces FEs at CFB Trenton and their responses solicited in one session. For the most part their responses were positive and the interface concept was seen as 'a step in the right direction.' It's potential contribution to 'de-snagging' aircraft systems was commented on favourably.

Focus group comments supported the use of the polar star diagrams and the state diagram graphics of the various aircraft systems. Some of the details of the mass-flow diagrams needed further explanation. Several FEs saw the value of the EID interface as a training aid to better provide systems-level information to operators - an area acknowledged to be in need of improved training (17).

In the context of designing for multi-operator systems, one interesting reaction to the moving map display was that "the Flight Engineer doesn't need to know that because he is not responsible for it!" Thus, the EID approach of providing all the information on a system that might be needed to diagnose an unanticipated problem was seen by potential users as affecting their roles and responsibilities.

6. DISCUSSION

At the time of the project, no other application was known of EID to aircraft interface design. This trial application was useful in showing that the principles of EID can be applied to aircraft systems, that EID information display concepts developed for other applications can be applied to aircraft systems, and that the level of effort required to implement an EID analysis of aircraft systems is not excessive. The review of the prototype by the FE focus group supported the conclusion that EID had produced an improved OMI and had identified a number of areas where the current system does not provide information that they need.

Given the age of the CC-130 OMI, any redesign involving the application of computer-driven displays should be an improvement. Thus the value-added by EID to the design of a modern aircraft remains undetermined by this project. It could be argued, for example, that a 'conventional' human engineering approach that included fault diagnosis among the FE's tasks should provide the much the same information as that identified by the EID analysis, particularly since the latter used display concepts which were taken from 'conventional' human engineering research. Other recommended approaches for OMI design include frameworks for analyzing display requirements which might produce results similar to those provided by the abstraction hierarchy. For example one of the authors was involved in the development by DCIEM of the OMI for a successful ship's machinery control system using Singleton's recommendation (18) that display systems provide information about: policies and objectives; alternatives and consequences, and; past, present, and future status.

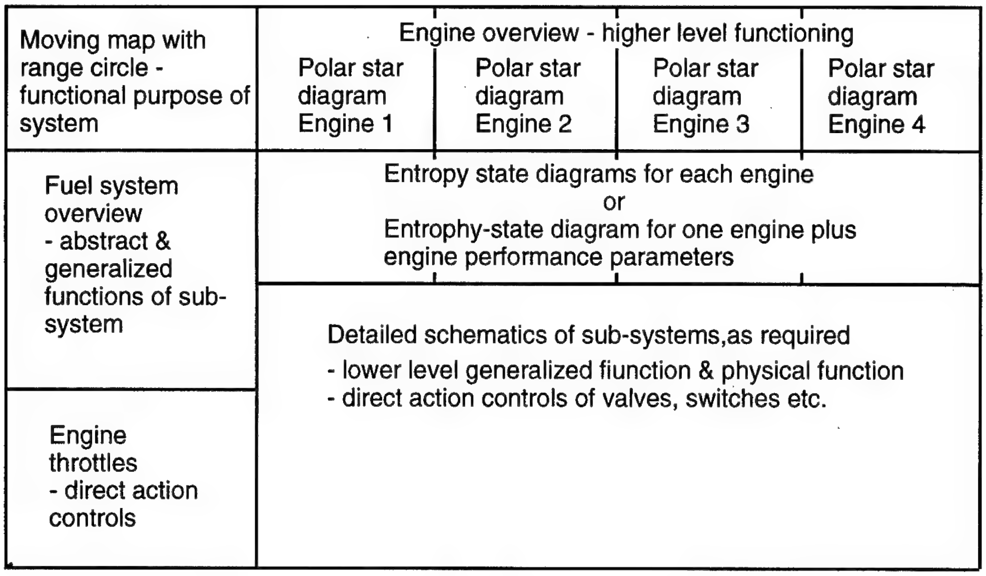


Figure 7: Schematic layout of the EID interface

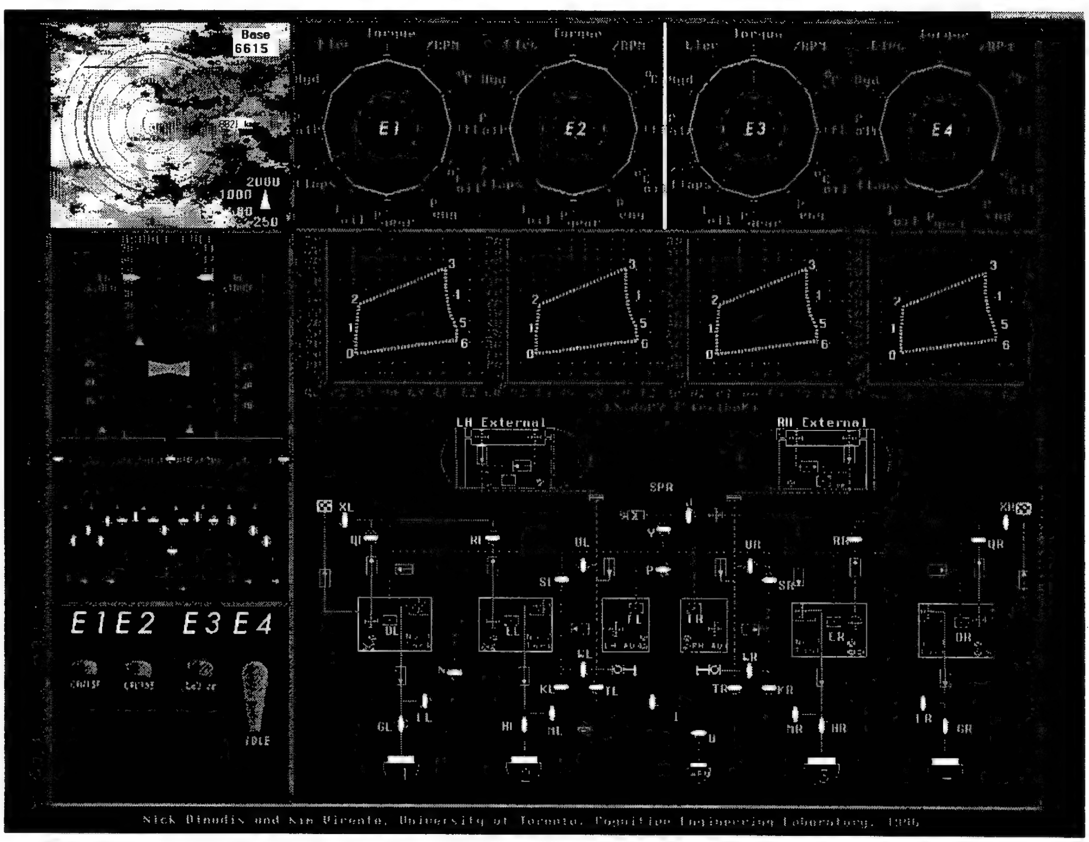


Figure 8: VAPS implementation of the EID interface (Engine 4 at idle, fuel system details selected)

The OMI produced using those guidelines included sub-system overview displays providing functional information, and selectable, successively more detailed displays of machinery status, structure and physical function.

Since comparative trials of these different approaches to human engineering are unlikely, claims of the superiority of one approach over another are best set aside in favour of identifying where EID can contribute to the human engineering design process. The prime contribution of EID in this exploratory application was in redirecting the focus of the OMI analysis to higher levels of system functioning. This is, in itself, of value; P.C. Schutte (personal communication, 1997) at NASA Flight Dynamics and Control Division has since run an experiment in which the diagnostic behaviour of pilots changed when knowledge-based information was provided. The EID matrix of analytical/ whole-part hierarchies identified information that is required to understand how each component, sub-system and system is functioning; those requirements would not have been identified by a design approach based strictly on providing state information using mimic diagrams.

In contrast to the OMI developed by EID, most of the system information provided by the current OMI is at the physical function/ component level (the lower right-hand corner of the analytical/ whole-part hierarchy matrix, Figures 2, 4). While it could be expected that an OMI based on 30 year old technology would focus information at that point of the matrix, it is not clear that such a focus would be avoided in the usual approach to human engineering that derives operator tasks from a third or fourth level of system functional analysis. To test this, a recent analysis of crew tasks for the CP-140 maritime patrol aircraft (which has the same engines as the CC-130) was examined. The analysis listed tasks such as:

- monitor and compare horsepower indicators
- monitor and compare turbine inlet temperatures
- monitor and compare RPM indicators
- monitor and compare fuel flow indicators
- monitor and compare oil pressure indicators, etc.

This list compares closely with the information provided by the polar-star diagrams developed for the EID interface of the CC-130 (Figure 6 and 8). This suggests that an OMI based on the task analysis would address the higher level functioning of sub-systems and units. Diagnosis of system problems was not included in the task analysis, so cannot be expected that any proposed OMI based on the task analysis would reflect diagnosis tasks. Further, an analysis of diagnostic tasks would not necessarily identify all of the operator's information requirements unless the task of diagnosing a particular problem was decomposed from the systems analysis. EID sidesteps this difficulty because it does not focus on operator's diagnostic tasks but on the system being controlled and the representation of the system functional state at a hierarchy of levels, from the overall system down to components.

A related contribution of EID that was observed in the exploratory application was that it provided an overall structure for the display of system information. The rules for implementing the SRK framework provided the basis for representing the aircraft systems information in ways which supported simple control behaviour (opening and closing valves), more complex behaviour involving several system components (sub-system state and mimic diagrams) and knowledge-based problem solving (entropy-state diagrams and mass-flow diagrams). Overall, in the aviation environment context, EID may be seen as making a specific contribution to designing for 'situational awareness.'

It could be argued that the various displays such as polar stars, maps, reservoir status and mass flow rates, and entropy state diagrams represent relevant 'metaphors' for the interface which can be used by designers without the need for an underlying theory, in the same way that others have suggested using the 'cockpit' as a metaphor for designing other control systems (19). On the other hand, the principles of EID provide a systematic basis for OMI design that can be taught, and experience from previous applications made an important contribution to the OMI concept. Thus EID provides a principled, rather than an ad-hoc approach to design. At the same time the project demonstrated the need to further develop principles for the translation of information requirements into a display system that is appropriate for a specific class of users. As with other human engineering formalisms, the principles of EID by themselves do not provide detailed guidelines for the representation of the visual form of the information identified by the analysis.

The project also showed that EID has implications for crew personnel and training issues. EID emphasizes designing for the three levels of SRK behaviour. However, not all operators are selected or trained to act at a knowledge-based level. This is increasingly the case as systems designers seek to reduce the training burden by trading it against automation and interface design. For example, most current multi-engine aircraft do not have a Flight Engineer and some current systems are not intended to be diagnosed by the operator but to be taken out of service until they can be properly diagnosed and serviced. The analysis of the CC-130 aircraft systems showed that such an approach would have to be matched by a very thorough analysis of sub-system and component interactions so that taking a faulty component out of service does not incur unexpected side effects. This returns the argument to the fundamental principle of EID, that designers cannot anticipate all events and that at some point an operator is going to have to diagnose problems and identify remedial action.

What is not clear is the extent to which an EID-based design affects the operator training burden. If an operator is not required to act at the knowledge-based level, then the operator training burden may be reduced. Given recent reports about reduced training and reduced levels of flight crew competence and systems knowledge (17; 20 & 21), this may not be the right approach. The focus group's reaction to the conceptual EID interface was that it would be a very useful training aid (an observation supported by other EID applications). This may justify a closer investigation of the potential contribution of EID to reducing training requirements and improving aircrew systems knowledge.

Of particular relevance to this symposium, the focus group's response to the EID interface showed that the design of information systems for one operator may have implications for the roles and functions of other crew members. Allocation of functions or tasks to operators is part of 'function allocation' (see Figure 1 & Refs. 22, 23) and may involve considerations of crew composition and rank. 'Function allocation' to crew members is not part of the EID approach because it assumes that functions have been allocated to operators by the design of the system and its components.

Thus, at present, EID does not address crew-operated systems. In this, EID is no different to most human engineering design models which focus on *a* human: machine system, rather than *multi* human machine systems. Design for collaborative crew performance poses a broader range of problems. Crew performance is dependent on processes which establish common goals and establish and maintain a common mental model (17). The design of crew-operated systems must take into account the allocation of functions to the various operators and the need to facilitate tasks associated with coordination, consultation, resolution of ambiguity, maintenance of awareness of system state, reversionary mode operation, training, crew performance monitoring and maintenance of alertness (24). Given the experience of the exploratory application reported in this paper, it is concluded that EID could make a contribution to many of those requirements if it is applied within a framework of designing for multi-operator systems.

7. REFERENCES

1. AGARD "Flight Vehicle Integration Panel Working Group 21 on Glass Cockpit Operational Effectiveness," AGARD-AR-349, April 1996.

2. Beevis, D., "Experience in the integration of human engineering effort with avionics system development." *in*: "The design, development and testing of complex avionics systems," AGARD-CP-417, April/May 1987, Paper 27.

3. NATO, "Application Of Human Engineering To Advanced Aircrew Systems," STANAG 3994, 1994, Brussels: NATO Military Association for Standardisation Aircraft Instruments Panel.

4. Stringer, F.S. (Ed.), "Optimisation of pilot capability and avionic system design," AGARD-AR-118. November 1978.

5. Fogel, L.J., "Biotechnology: concepts and applications," Englewood Cliffs, N.J., Prentice-Hall, 1963, pp 599-684.

6. Stokes, A.F. and Wickens, C.D. "Aviation displays," *in* E.L. Weiner and D.C. Nagel (Eds.) "Human factors in aviation," 1988, San Diego, Academic Press, Chapter 12.

7. Vicente, K.J. and Rasmussen, J., "Ecological interface design: theoretical foundations," IEEE Transactions on Systems, Man, and Cybernetics, 22, 4, 1992, pp 589-606.

8. Bisantz, A.M. and Vicente, K.J. "Making the abstraction hierarchy concrete." Int. J. Human-Computer Studies, 40, 1994, pp 83-117.

9. Rasmussen, J., "The role of hierarchical knowledge representation in decision making and system management," IEEE Transactions on Systems, Man, and Cybernetics. SMC-15, (2), 1985, pp 234-243.

10. Vicente, K. J., "Improving dynamic decision making in complex systems through ecological interface design: A research overview," System Dynamics Review, 12, 1996, pp 251-279.

11. Rasmussen, J., "Skills, rules, knowledge: Signals, signs, and symbols and other distinctions in human performance models," IEEE Trans. on Systems, Man, and Cybernetics. SMC-13, 1983, pp 257-267.

12. Vikmanis, M.M., "Advances in workload measurement for cockpit design evaluations," *in* "The man-machine interface in tactical aircraft design and combat automation," AGARD-CP-425, July 1988, Paper 10.

13. Reising, J.M., and Emerson, T.J., "Color display formats: a revolution in cockpit design," *in* "Proceedings of the Conference on Advanced Avionics and the Military Aircraft Man/Machine Interface," AGARD-CP-329, April 1982, Paper 6.

14. Turek, R.O. and Greenstein, J.S., "Stars, polygons, and clusters," *in* "Information management and decision making in advanced airborne weapon systems," AGARD-CP-414, April 1986,. Paper 8.

15. Mumaw, R.J., Woods, D.D. and Eastman, M.C. (1992). "Interim report on techniques and principles for computer-based display of data." (STC Report 92-1SJ4-CHICR-R1). Pittsburgh, PA: Westinghouse Science & Technology Center.

16. Dinadis, N. and Vicente, K.J., "Application of Ecological Interface Design to aviation." DCIEM Contract Report No. 97-CR-02, 1996, Toronto, University of Toronto, Cognitive Engineering Laboratory.

17. Hendy, K.C., Thompson, M.M., Fraser, W.D., Jamieson, D.W., Comeau, P., Mack, C.I., Paul, M.A., and Brooks, C.J., "Human factors of CC-130 operations — Volume 1: Final report of the DCIEM/ATG study team" DCIEM 98-R-14, 1998, Toronto, Defence and Civil Institute of Environmental Medicine.

18. Singleton, W.T. (1972). "General theory of presentation of information," *in*: R.K. Bernotat and K.-P. Gärtner (Eds.) "Displays and controls: proceedings of an Advanced Study Institute," Amsterdam, Swets & Zeitlinger. pp 75-81.

19. Colgan, L., Spence, R., and Rankin, P., "The cockpit metaphor." Behaviour & Information Technology, 14 (4), 1995, pp 251-263.

20. "IATA says lack of pilot competence is biggest safety problem," Flight International, February 25, 1998, p. 4.

21. "The wrong stuff." Flight International, February 28, 1998, p. 3.

22. Beevis, D. Essens, P. and Schuffel, H. (Eds.), "Improving function allocation for integrated systems design." Wright-Patterson Air Force Base, Ohio, Crew Systems Ergonomics Information Analysis Center, 1994.

23. Fallon, E.F., Bannon, L. and McCarthy, J. (Eds.), "ALLFN'97, Revisiting the allocation of functions issue: new perspectives," 1997, University of Louisville, KY, International Ergonomics Association (IEA) Press.

24. Beevis, D., "Human functions and system functions." *in*: D. Beevis, P. Essens, & H. Schuffel (Eds.). "Improving function allocation for integrated systems design," 1994, Wright-Patterson Air Force Base, Ohio, Crew Systems Ergonomics Information Analysis Center.

Collaborative Crew Issues in a Future Reconnaissance Land Vehicle

M Thody and I F Ross
British Aerospace, MA&A
York House, PO Box 87
Farnborough Aerospace Centre
Farnborough, Hampshire
GU14 6YU, UK.

1. SUMMARY
TRACER/FSCS is an international collaborative programme to develop an armoured reconnaissance land system for the British and US Military and is set to enter the project definition phase from which a design solution will be proposed. The vehicle will be small and stealthy, designed to carry out surveillance and intelligence acquisition on tomorrow's battlefield, and be equipped with state-of-the-art sensors, communications, battlefield information, and weapon systems. The TRACER/FSCS crew, of three soldiers, will be required to execute a complex and demanding role through the effective and efficient operation of these complex systems. Overall system performance will necessitate an effective collaborative team performance from three individual soldiers forming an integrated crew. The TRACER/FSCS programme provides a considerable HFI challenge and numerous questions have been raised regarding crew collaboration that will be addressed during development. This paper highlights seven key areas for discussion and details the subsequent conclusions.

2. LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AFV	Armoured Fighting Vehicle
BAe	British Aerospace
CVR(T)	Combat Vehicle Reconnaissance (Tracked)
DAS	Defensive Aid Suite
FSCS	Future Scout and Cavalry System
HF	Human Factors [Ergonomics]
HFI	Human Factors Integration
II	Image Intensifier
IKBS	Intelligent Knowledge Base System
ISA	Instantaneous Self Assessment
NASA	National Aeronautics and Space Administration
SME	Subject Matter Expert
SMI	Soldier Machine Interface
TI	Thermal Imager
TLX	Task Load Index
TRACER	Tactical Reconnaissance Armoured Combat Equipment Requirement
UK	United Kingdom
US	United States [of America]

3. INTRODUCTION
TRACER/FSCS is an international collaborative programme to develop an armoured reconnaissance land system for the British and US Military. The programme is soon to enter the project definition phase where a joint venture company, SIKA International, formed by an international consortium of which BAe is prime systems integrator, will propose a design solution.

The role of TRACER/FSCS will be to provide close reconnaissance for the battle group in all phases of war, in particular giving the battle group commander timely and accurate information by day and night in all weather conditions. To achieve this role, TRACER/FSCS will be a small, stealthy armoured fighting vehicle designed to carry out surveillance and intelligence acquisition on a digitised battlefield. Equipped with II and TI systems, a battlefield radar (both systems mounted on a retractable mast), a cannon and a comprehensive navigation, command and control, communication and DAS system. All of which are to be operated effectively and efficiently by a crew of three on missions deep in enemy territory.

It has long been recognised that the reconnaissance role is an extremely demanding to achieve and one on which overall mission success is dependant. This was illustrated eloquently by Solzhenitzyn[1] when he wrote "The need for rapid action was now acute. Within the space of a few minutes he had to grasp the lay-out of the ground, determine the enemy's positions and his own, select a defensive line to be occupied by the Ladoga battalions, agree with the gunners on a common observation post, pay out the telephone wires, and shoot in the guns by registration fire on fixed targets. If within those few minutes mistakes were made in organising, selecting or dispatching, and the orders were given in the wrong sequence or incorrectly then they could not be put right within the next half hour, and if in that half hour the enemy attacked or started firing, then the men's keenness, their good telephone communications and their sixty rounds per barrel would all be useless; they would simply have to run for it.". Technology may have moved on since 1914, but the philosophy of reconnaissance remains unchanged. Therefore by necessity, the crew of TRACER/FSCS will perform a multifarious and demanding role using a number of different components that will form an integrated but complex system. To achieve mission success, each soldier will be a key player within a cohesive three person team that will affect collaborative crew performance under exceptionally demanding operational requirements.

The development of TRACER/FSCS provides a considerable HFI challenge and, as suggested by Nicholson and Horner (1991), the successful incorporation of HF requirements to the design of complex military systems requires a systematic and accountable input throughout all stages of development. With this in mind, consideration of the various HF issues required to deliver an *Ergonomic* solution must be made from the start of project definition.

[1] English translation by Glenny, (1972).

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

Key HF issues have been identified; these will be addressed during the development programme of TRACER/FSCS; these issues currently are in the form of questions rather than answers. This paper outlines a subset of those key issues that are related to collaborative crew performance and discusses the impact of implementing various design solutions. Conclusions have been drawn indicating how these issues could be addressed formally during development and what should be considered in potential trade-off analyses.

4. COLLABORATIVE CREW ISSUES

The key HF issues related to collaborative crew performance that have been identified for discussion here are:

- Crew Work Station Design
- Sharing Controls and Displays
- Task Distribution
- Maintaining Situation Awareness
- Crew Performance and Workload

4.1 Crew Work Station Design

The design of the crew work station to facilitate crew collaboration, for example simultaneous operation of controls and displays, cueing of displays and executive control, is essential to the efficient and effective execution of the reconnaissance role from within the TRACER/FSCS system.

It is safe to assume that the size and profile, dictated by the intended role of TRACER/FSCS, will be prime drivers in the design process of the vehicle. This will have a direct impact on the internal volume and, hence, the space envelope available for the crew work stations. This is typical of AFV design, where work stations are notoriously cramped and cluttered and invariably uncomfortable. Historically, these features have been exacerbated by the agricultural nature of the equipment installed. The advanced technological equipment fit, to meet the requirements of TRACER/FSCS, will be far from agricultural; this presents the opportunity to provide the user with a more comfortable and less cluttered working environment. However, to provide the soldiers with access to the numerous systems, the work stations will require a comprehensive SMI which is likely to occupy significant interior real estate. As a consequence, the trend of limited crew work space in AFVs will probably continue.

The limited crew work space envelope does, however, provide the opportunity for locating two work stations in close proximity which lends itself to collaborative crew interaction. Soldiers working in this close proximity has its advantages, but also presents some potential difficulties that should be considered. Examples of the positive and negative attributes were reported, respectively, by Harvey (1994). They were: increased team cohesion due to the immediate feedback and ease of interaction between the crew; improved Command, Control and Communication aspects of crew performance due to the physically open nature of the crew space; exaggerated intensity of the partnership due to lack of privacy; and continual awareness of the other crew member in the peripheral visual field preventing the venting of negative emotions without the other crew member noticing.

There are other issues arising from the drive towards enabling production of a small and stealthy vehicle by reducing the crew work space to a minimum. For example, the use of multi-functional control panels provide the soldier with access to various systems and their functions via the same set of switches. Control panel dimensions can, obviously, be reduced if they have less switches. However, Tyler, Gosling and Barber (1992) stated that speed to initiate the desired interaction decreased as the number of hierarchical levels increased. The MoD HF guidelines, DEF-STAN 0025, also suggest the use of broad shallow hierarchical structures. The trade-off between SMI dimensions and usability will be considered carefully through a comprehensive rapid prototyping programme.

Reclining soldiers in their work stations is one method of reducing the interior floor to ceiling height which has a direct impact on the external profile of the vehicle. There is, however, a limit to the angle of recline that soldiers can tolerate before their performance is likely to suffer. An experimental study concluded that a seated posture of 60° recline does not afford optimum vigilance performance (Thody, Gregg and Edwards, 1993). This is so, particularly, if the SMI is mounted in the plane perpendicular to the floor as postural loads in the soldier's back and neck increase causing significant discomfort. However, one important aspect of the reclined seat is that it offers better quality of sleep than an upright sitting posture (Nicholson and Stone, 1987). This should be considered because if soldiers could get some good quality sleep, even for short periods (between 2 to 4 hours), they may be able to retain a stable work rate and level of performance throughout a typical battlefield mission. It is important, for successful crew collaboration, that all the crew members are able to achieve and maintain a stable work rate and level of performance; if not, it is probable that workload will be distributed unevenly leading to a reduced probability of mission success.

Further to the provision of a comfortable and efficient work space envelope, consideration must be given to the functions provided at each crew work station. Fundamental to this is the question: Should each crew work station be tailored to an individual crew role or, should they have similar SMI to provide all crew members with a common working environment?

Historically, due to the mechanical construction of the vehicle's systems, crew work stations were tailored to specific tasks. For example, the driving controls were linked mechanically to the engine and transmission, the weapon systems were loaded manually and controlled via mechanical hand wheels, and the optical vision devices were simply roof mounted periscopes. These limitations made duplication of functionality practically impossible from both an engineering and cost perspective. To operate such a vehicle, a crew of three role specific members was required, namely Commander, Gunner, and Driver. Task sharing and balanced workload distribution is extremely difficult to achieve with this crew configuration with each member experiencing periods of work underload and overload during a mission. Also, if a particular crew member is injured or killed then their tasks go unaccomplished; for example, no driver means no vehicle mobility and no gunner means no fire power. However inflexible this is, individual soldiers know exactly what tasks they are responsible for within a clearly defined job.

The introduction of *fly by wire* technology enables mechanical linkages to be replaced by micro-switches and actuators which has the potential of duplicating functionality at each work station within the vehicle. These generic work stations provide the potential for even distribution of tasks and workload and the facilities to take on another's tasks in the event of their incapacitation. Some tasks will still be executed more efficiently from specific work stations, but this will be related to their location within the vehicle. For example, driving will be easier from a work station in the vehicle hull than from one in a turret because disorientation will result if the driver does not face the direction of travel. With this in mind, generic work stations do not necessarily need to be identical in appearance or layout; but they should provide the same functionality, nonetheless.

A number of factors have been discussed that have a direct impact on the design and implementation of work stations for TRACER/FSCS. Clearly, there is no simple answer as to the best design solution. However, the components critical to the inevitable trade-off analyses have been identified for consideration. Achieving the right balance between these factors, for example common functionality, common interface design, role specific tuning, crew work space, and vehicle dimensions will require extensive consideration and rigorous evaluation.

4.2 Sharing Controls and Displays

Following on from general crew work station design, the ability to share controls and displays has the potential for enhancing the performance achievable from collaborative crew interaction. The flexibility offered by technology to enable different sensors and digital images, for example a digital map, to be displayed on any screen within the vehicle either separately or mixed has enormous benefits for crew collaboration. However, the crew collaboration is not limited to a single approach because the flexibility provides the crew members with the facility to view the same image or different images simultaneously. This allows the crew to work as an integrated team in, broadly speaking, three different ways.

Firstly, the crew members can work closely together with their collective attentions focused on the same image which will encourage discussion. This is beneficial for comprehending and confirming information which facilitates informed decision making which will lead to the recognition and agreement of common task goals. This has the powerful effect of improving team cohesion through the reduction of feelings of individual isolation.

Secondly, the team can work on separate parts of the same task, for example each looking at a different sensor image of the same scene would provide a comprehensive and complementary information set of the tactical picture. This approach also enables informed decision making and strategic planning. Further, it has the potential for reducing individuals' workload through task sharing and distribution.

Thirdly, the team members concentrate on their individual set of tasks that, collectively, enable mission success but with the facility to provide help or support to each other immediately. This key benefit, offered by the ability to share displays and controls, increases team cohesion through mutual support. It also enables rapid task redistribution because all the crew have an awareness of what each other is doing, the information they are receiving and processing, and how it relates to the overall mission.

To optimise the sharing of controls and displays, there are some physical interaction issues to consider. These include: sizes of, positions of and viewing angles to displays; how to prevent selection of controls that cause task conflict; position of controls within multiple reach envelopes; and use of controls with left and right hands.

The sharing of some images may be more appropriate if actually displayed on a single screen. Typically, direct interaction of pointing a finger or pen at key features on the digital map would aid discussion and comprehension of tactical information. This would be possible only if the display surface was positioned within the overlapped visual fields and reach envelopes of the collaborating crew members. Similarly, access to shared controls would have to be within overlapping reach envelopes.

Shared controls should be equally usable with the right or left hand and be uniformly shaped. They should not be tailored to fit the contours of just the left or the right hand because, if sat side by side, the soldier located on the left will operate the control right handed and the one sitting on the right will use their left hand. There are also implications for groups of associated controls, for example a tracker ball could be controlled either left or right handed but the associated selection switch (usually activated by the operator's thumb) might have to be duplicated to accommodate both left and right handed interaction (Figure 1).

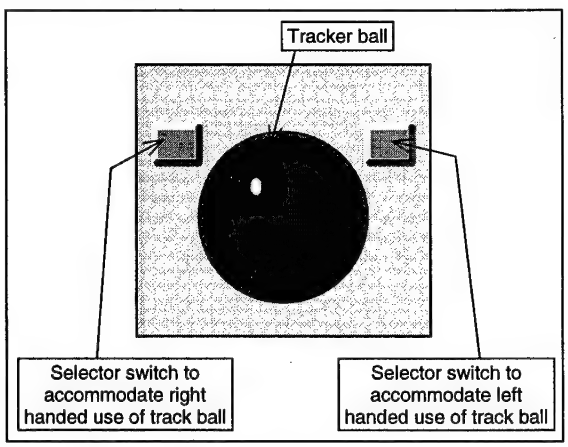


Figure 1. Tracker ball with duplicate selector switches.

Preventing interaction, inadvertent or not, of shared controls that cause task conflict is an important issue. Critical functions, such as weapon firing or vehicle steering, should be possible only from one crew work station at a time. Exclusive selection interlocks should be implemented together with override priorities that may be dependent on soldiers' rank, role, or work station location.

The issues pertinent to shared controls and displays, as discussed above, will be addressed throughout the design programme. This will be achieved through systematic analyses of prototype SMIs using rapid prototyping techniques and evaluating virtual representations of the

vehicle work stations with a bespoke computer aided human modelling tool.

4.3 Task Distribution
Allocation of functions between humans and automated systems used to be a relatively easy process because there was a clear distinction between what humans were good at and what machines were good at. However, with the advent of AI, IKBS and cognitive decision aiding systems this distinction has become somewhat blurred. So far, the allocation of function has tended to be done on the basis of what AI/IKBS systems are available and not necessarily where individual operator or combined crew workload is high or where an automated system could improve overall system performance. This results in dubious design traceability, to say the least, and demands that the design trade-off process needs to be undertaken and duly recorded rigorously. The consequences of implementing AI and IKBS systems without a rigorous HF design strategy will be disjointed function sets for operators that will increase workload which in turn will reduce their performance and that of the crew and overall system.

Historically, the strategy to address the issues of allocation of function between humans and systems can be described has having had a number of iterative phases. Initially, a set of typical mission scenarios was identified and generated that covered the full scope of primary activities to meet the role and mission requirements of the system. From this, a set of generic mission phases was derived that could be assembled to form any mission expected of the particular system and specified the most demanding 'forcing' mission modes and segments associated with each phase. Following a review of the potential capabilities of the human, the system and equipment components, a list of crew and system tasks for each segment was determined.
Finally, the functional requirements together with the degree of automation acceptable and appropriate would be described. Each system function and task would then be allocated to human operator, automated system (for example AI, IKBS, or cognitive decision aid) or a combination of both.

It is recognised that this system has its limitations in that it does not specify the allocation of function criteria. Furthermore, it does not take into account any specifically defined operator capability, nor operator or crew workload, and does not include other pertinent criteria such as those for manpower personnel trade-off and life cycle costings.

For TRACER/FSCS the following improvements should be made. Consideration must be given to the fact that TRACER/FSCS will have a crew of differing capabilities, for example, in terms of experience, aptitude and rank. Therefore, the role of the operator should be defined to as detailed a level as possible based on the target audience description. Timeline data from operational analysis studies should also be considered with a view to producing workload predictions. To perform a systematic trade-off between human operator and system operation and the training, manpower and personnel issues, the design criteria and allocation of function rules must be agreed and stated explicitly. These criteria and rules should be produced to include the necessary performance figures by which the trade-off analyses are governed.

This crucial issue must be addressed and should include the identification of system functions and their suitability for the application of AI/IKBS and cognitive decision aiding techniques. The findings from such a study will be invaluable to focus the necessary attention toward developing and improving the process of allocating functions in a systematic manner.

4.4 Maintaining Situation Awareness
Possessing good situation awareness is critical for the individual soldiers and the crew, as a cohesive unit, to achieve mission success. The requirement for providing operators, of complex systems, the facilities with which they can build and maintain their SA is understood and accepted. For example, Taylor and Selcon (1990) stated that SA is needed for interface design to proceed beyond workload reduction towards providing solutions to mission problems. Indeed, the role of armoured reconnaissance is to be the *eyes and ears* of the senior battle directors, to provide them the wherewithal to build and maintain a comprehensive SA. The workload required to maintain the flow of timely and relevant information will, alone, be demanding. Therefore, the quick and easy maintenance of their own SA, without the need for excessive additional workload demands, will be extremely beneficial to the crew of TRACER/FSCS.

Before TRACER/FSCS, SA was achieved by the vehicle crew having direct contact with the environment and absorbing raw information through their senses. Typically, the information sourced directly would be; three dimensional view of the terrain (either through naked eye or optical vision device, e.g. binoculars), three dimensional sound, vehicle position and motion relative to surroundings, and environmental characteristics. This information enables the building and maintenance of a detailed and accurate SA, but one which is limited to the capabilities of detection and comprehension by the human senses. To achieve the maximum from the senses, soldiers operate head-out which involves putting their personal sensor package, i.e. the head, outside the protection envelope of the vehicle through a hatch opening in the roof. Obviously, exposing one's head to the battlefield environment has potentially serious risks attached. However, in the absence of current technological advances, taking these risks was necessary to achieve suitable SA.

The technology available for TRACER/FSCS has provided the opportunity to present external information indirectly to the vehicle crew via various electronic sensors. As well as enabling the crew to operate more safely, the technology has the potential for increasing the coverage of external information gathering and pushing the SA envelope beyond that of the immediate locale. To realise this potential however, the sensors' capabilities, display format design, information presentation, ability to share direct and indirect vision devices, will have to be integrated with care. The indirect information will be presented in such a way as to complement the soldiers' SA building capabilities. This may include providing additional information to compensate for that which has been lost as a result of the physical separation from the external environment.

It is anticipated that the crew of TRACER/FSCS will build and maintain their SA primarily from the information presented in the form of a digital map on a colour display. The digital map format lends itself to

providing additional information that, if kept up dated in real time or with minimal lag, will aid the crew to maintain their SA. Critical to the maintenance of SA, is knowing own vehicle position and heading, but this is only half the story. Essential to efficient and effective navigation is knowing own vehicle position relative to intended destination and to potential routes to achieve that destination. This information is best provided by presenting a comprehensive set of symbology and text that overlays the map image. The map overlay would, for example, include: an icon indicating own vehicle position and orientation driven by GPS; a textual read out line including time, heading, and heading to next waypoint; icons to indicate waypoints for route planning and following; and icons representing friendly and enemy forces.

The ability to display the digital map in three dimensions would enable further enhancement providing the crew, for example, information about terrain profile and intervisibility. This information would form part of a comprehensive and integrated AI/IKBS that the crew would access to aid SA and enable informed mission planning.

The use of complex sensor systems presents the potential for individual or crew disorientation. As mentioned earlier, TRACER/FSCS will be equipped with various visual sensors capable of viewing in any direction independent of each other, the primary weapon system, and the vehicle. Each crew member must be provided with meaningful and unambiguous cues to inform them of the various directions that the active sensors, weapons and vehicle are pointing. Value will also be gained from knowing how the visual coverage of the active sensors relates to vehicle position and surrounding terrain. This information could be included as part of the own vehicle icon and displayed on the digital map. There is, however, a requirement to exercise caution in the design and implementation of what will inherently be a complex symbol. For example, the likely components of an icon to represent own vehicle position and its sensors is shown below (Figure 2).

As mentioned earlier, the role of armoured reconnaissance and the ability to maintain SA also requires use of the ears, but they like the eyes will be separated from the external environment as a result of operating within the closed down TRACER/FSCS. The provision of real-time three dimensional sound could be made to the crew members via an array of microphones, fitted to the outside of the vehicle, and their headphones. Other SA building information that is available to and used by the head out operator should be provided via sensor systems or IKBS and displayed at the crew work stations. For example, air temperature, wind speed and direction, cloud cover, and visibility.

As discussed, the soldiers of the crew must absorb and assimilate huge quantities of diverse information, some of which sounds trivial when considered in isolation, both consciously and subconsciously through all their senses to enable them to build and maintain a comprehensive SA. Removing soldiers from direct contact with their immediate external environment will have its consequences, but with careful and rigorous consideration they should be minimised mostly and in some cases the information may be enhanced.

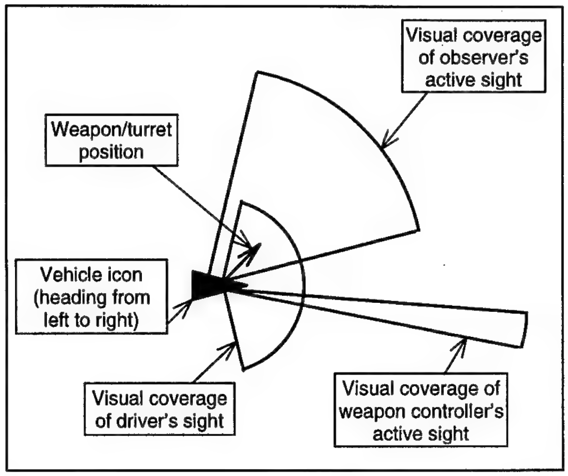


Figure 2: Example of icon included in map overlay.

4.5 Crew Performance and Workload

The overall performance of TRACER/FSCS will depend on the crew being able to operate, as an integrated team, in an optimum manner. This in turn will depend on the individual crew members being able to perform the individual tasks that make up the crew functions. For the crew to operate effectively, they must be neither over or under loaded as a crew or as individual operators. Ensuring an acceptable level of crew workload under all operational circumstances is a TRACER/FSCS design requirement.

Predicting operator workload during the early stages of design and development is essential. The cost of redesign increases with design maturity and, therefore, changes must be identified and incorporated early. An effective and integrated design process, where crew workload is an identified and accepted design criteria, is essential together with the appropriate Systems Engineering and HF tools and techniques including rapid prototyping of the SMI. The role of Subject Matter Experts is vital in providing in-depth operational knowledge of both operator tasks and operator capabilities as design input and criteria, during design and development. The SMEs are an integral part of the design team and will be involved in the early and frequent informal assessment of the design during early development, prior to formal assessment.

The assessment of crew performance and workload is not simply the summing of individual operators' performance and workload and as such can not be quantified by the addition of the performance and workload achieved by individuals. Assessment must be at a crew level and the techniques must be simple to apply and robust. This results in measurement at a gross level and interpretation by comparison with other existing and understood systems.

TRACER/FSCS is a considerable leap in terms of systems' complexity when compared to the UK's existing reconnaissance vehicle CVR(T) and comparison of operator tasks can only be made at a top level. If a comparison is to be made, the top level comparable functions must be identified and crew performance and workload assessed. This must be carried out using the same techniques and methodologies and preferably by

the same engineers to avoid bias. Other questions arise which must be addressed if the comparison is to be valid, for example, is the same crew used for both assessments on the existing system and the developed system? The crew may be experienced users of CVR(T) whose task performance has been enhanced by many hours of operation. How can such proficiency be attained on a new system that may only be in prototype form? These questions and their resulting issues are, obviously, typical for the domain of complex systems design and development where new systems are quite different from those that they replace. However, the answers are not so clear cut and are addressed heuristically through the experience of the design and development team. In this respect, the key personnel on the team are the programme SMEs who will have an intimate knowledge of the new system through close involvement during development. It is recognised that this method is less than ideal and, although practised widely, may not be wholly acceptable to the assessment authority.

The techniques used to assess crew performance and workload need to be appropriate, robust, consistent and reliable. They must be quick and easy to apply by engineers and incur no large cost overhead that must be borne by the project. Perhaps not surprisingly, due to the inherent difficulties of developing and implementing such rigorous techniques, none have been found to exist. Subjective techniques such as ISA or NASA TLX are appropriate for individual operator assessment but not for crew assessment. For assessment of crew performance a further technique must be used. First, crew functions and their appropriate measures of performance must be identified with the aid of SMEs. Preferably, these functions should also be carried out in the previous system and therefore allow comparisons to be made, for example, target detection times and target engagement times. The functions are carried out on both systems, the performance measured and comparisons made. It would be this easy if not for all the associated, and mostly uncontrollable, difficulties and uncertainties that make any result open to interpretation.

5. CONCLUSIONS

As suggested at the beginning of this paper and as a result of the following discussion, the development and implementation of a successful design solution to TRACER/FSCS quite clearly provides a considerable HFI challenge. There are no clear cut answers provided here, but the paper has identified the components critical to the inevitable trade-off analyses that will need to be undertaken in a systematic and fully accountable manner throughout all stages of development.

The SIKA consortium have an extensive and specialist Human Factors experience that has previously been successfully applied within the domain of complex military systems and is, therefore, well placed to meet the challenge of developing a successful design solution to TRACER/FSCS.

6. REFERENCES

1. Glenny, M., "August 1914", published by The Bodley Head, 1972. (English translation of an original text by Solzhenitzyn, A).

2. Nicholson, L.M. and Horner, S.R.L., "Accountability of Human Factors in the Design of Complex Military Systems", in "Ergonomics - Design for Performance", Proceedings of the Ergonomics Society's 1991 Annual Conference, April 1991, pp 112-117.

3. Harvey, J.S., "Assessment of Team Cohesion and Team Working for the VERDI-2 System and the Scimitar Vehicle During Trials Conducted on SPTA July 1994", Annex E-DRA/FV&S3/CR94/031/1.0, November 1994.

4. Tyler, S., Gosling, P. and Barber, P., "Menu Depth vs Menu Breadth in Tank Control Panel Design", in "Ergonomics for Industry", Proceedings of the Ergonomics Society's 1992 Annual Conference, April 1992, pp 242-248.

5. Ministry of Defence., "Human Factors for Designers of Equipment. Part 13: Human Computer Interaction", DEF STAN 00-25 (Part 13)/Issue 1, May 1997.

6. Thody, M., Gregg, V.H. and Edwards, R.J., "Reclined Sitting Postures: Their Effect on Human Performance of a Vigilance Task", in "Ergonomics and Energy", Proceedings of the Ergonomics Society's 1993 Annual Conference, April 1993, pp 28-33.

7. Nicholson, A.N. and Stone, B.M., "Influence of Back Angle on the Quality of Sleep in Seats", Ergonomics, Vol.30.7, July 1987, pp 1033-1041.

8. Taylor, R.M. and Selcon, S.J., "Understanding Situation Awareness", in "Ergonomics Setting Standards for the '90's'", Proceedings of the Ergonomics Society's 1990 Annual Conference, April 1990, pp 105-111.

Collaboration in Complex Medical Systems

Yan Xiao and Colin F. Mackenzie

University of Maryland School of Medicine

10 S. Pine St., MSTF 534, Baltimore, MD 21201, USA

Email: yxiao@umaryland.edu URL: www.hfrp.ab.umd.edu

SUMMARY

Improving our understanding of collaborative work in complex environments has the potential for developing effective supporting technologies, personnel training paradigms, and design principles for multi-crew workplaces. Using a sophisticated audio-video-data acquisition system and a corresponding analysis system, the researchers at University of Maryland have been able to study in detail team performance during real trauma patient resuscitation. The first study reported here was on coordination mechanisms and on characteristics of coordination breakdowns. One of the key findings was that implicit communications were an important coordination mechanism (e.g. through the use of shared workspace and event space). The second study was on the sources of uncertainty during resuscitation. Although incoming trauma patients' status is inherently uncertain, the findings suggest that much of the uncertainty felt by care providers was related to communication and coordination. These two studies demonstrate the value of and the need for creating a real-life laboratory for studying team performance with the use of comprehensive and integrated data acquisition and analysis tools.

INTRODUCTION

Working together has been an inseparable facet in all human activities. The importance of understanding coordination grew in recent years because design of technical systems is more and more dependent on our conception of how people work together. Ever more powerful and accessible information technology tools are available to help people work together. Furthermore, it also became the realization that training people to work together could also prevent many of the failures and accidents in military and civilian systems.

Medical care provision is one of the primary examples of collaborative efforts over time at various locations by multi-disciplinary teams. It is both a fertile ground for studies of collaboration *in situ* as well as a domain which can benefit greatly from a better understanding of collaboration. The domain of trauma patient resuscitation can be characterized as involving:

- high risk
- severe time pressure
- activities from multiple, highly experienced specialists
- many unknowns about the patient

Typical sequence of events is:

1. activation of emergency medical services
2. first aid and field treatment
3. triage and transport, and
4. emergency and definitive treatment in a dedicated facility.

Collaborative efforts occur along this sequence of events:

- radio dispatch
- patient triage
- on-line consultation between field care providers and physicians
- emergency medical services resource management
- sharing of information among the clinicians preparing for the incoming patient
- execution of resuscitation plans within a compressed time window by a multi-disciplinary team once the patient arrives, and
- definitive care by a surgical and anesthesiolgical care team in the operating room.

Seven years ago, the Shock Trauma Center at University of Maryland established a video based data acquisition system for studying team performance during trauma patient resuscitation and anesthesia. When trauma patients were brought to the Center through Medevac helicopters or ground ambulances, their initial assessment/resuscitation and subsequent surgery could be videotaped and later reviewed through an integrated, comprehensive system. A library of multimedia records of trauma patient resuscitation and anesthesia was accumulated. In the library there were:

- patient admission records
- videotapes
- commentaries by participant and neutral subject matter experts
- patient laboratory findings, and
- discharge summaries

A unique feature of the video recordings was that patient vital signs (eg heart rate and blood pressures) were overlaid on top of the video image, thus making it easier for reviewers to determine patient status and the objectives relevant to the resuscitation (Figure 1).

In this paper, we will describe two studies of team coordination based on video analysis. We will conclude with a taxonomy of collaboration in terms of temporal scales and implications for system designs and training.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

Figure 1. A sample image from a videotaped case. The vital signs of the patient were displayed across the top: heart rate (HR) = 116, systolic blood pressure (SBP) = 117, mean blood pressure (MBP) = 66, diastolic blood pressure (DBP) = 51, end-tidal CO_2 = 28%, pulse oximetry = 99%. The case was in an operating room; the image was taken just after the anesthesia care providers finished anesthesia induction and airway management and just before the surgical crew was ready to operate.

VIDEO ANALYSIS METHODS

Over a period of four years, more than 120 trauma resuscitation and anesthesia cases were recorded using the data acquisition system. Major components of our video analysis included:

1. audio commentary by subject matter experts
2. coding of verbal communications
3. event flow analysis of selected cases
4. performance evaluation using a normative task model.

Reviews by subject matter experts provided insight into the potential cognitive processes involved in making diagnoses, plans, and decisions, and have become an important data source themselves.

Of special interest to the topic of collaboration is event flow analysis, which is a process of constructing detailed event flows with hypotheses or theories of what might contribute to the underlying cognitive activities. One example of results of such an analysis is given in Figure 2, which described a segment of a videotaped case. To facilitate video analysis, a system was developed which linked video with transcriptions of verbal communications, coding of events, and patient vital signs (Figure 3).

STUDY 1. TEAM COORDINATION AND BREAKDOWNS

This study was driven by the fundamental question of how it was possible for team members to function so smoothly most of the time with little apparent effort spent on coordination. The analysis taken to address this question was qualitative, aimed to:

1. categorize ways in which team coordination was achieved, and
2. hypothesize the nature of breakdowns in team coordination.

Three types of critical incidents were analyzed: decision points, high workload periods, and apparent problems in team coordination.

The findings related to coordination strategies that were used by resuscitation teams are reported in two separate areas: task coordination, or the distribution and delegation of tasks; and information flow, or the passage of information regarding patient status and contingency plans.

Task Coordination

During the course of resuscitating a trauma patient, many physical tasks were performed. Some of them had to be coordinated among team members within a crew or across crews. This was so either because the tasks needed synchronous effort from multiple people (e.g., lifting the patient), or because the tasks relied on preconditions (e.g., suctioning equipment must be ready before usage), or because multiple tasks need to be accomplished within a short period of time (e.g., establish the airway and restore blood circulation).

Several forms of non-communication task-coordination activities were noted in video analysis. Four of them are listed below.

Following the protocols. Established practices (sometimes codified as protocols, such as the Advanced Cardiac Life Support protocol), specify task distributions and priorities, immediate goals, and problems to be treated. The tasks to be done by each team member are clear. Without much communication, in almost every case, the surgical, anesthesia, and nursing crews commence their activities after the patient arrived. We observed clear task distributions among the crews in resuscitation teams at the beginning of each patient admission, despite the uncertainty about the patient's status.

Following the leader. Team members determined what they should do by monitoring the leader. The activities of the team leader can be viewed in some sense as the "medium" through which the team leader passed information (such as instructions) to the rest of the team. If not occupied, we observed that team members tended to follow the attention foci of team leaders. Needed materials or help were provided often without explicit solicitation.

Anticipation. The team members were also found to provide unsolicited assistance through the anticipation of the team leader's response to the patient's physiological events. A gagging sound, in one case, led an assistant to offer a suctioning catheter in anticipation that the patient would vomit soon and the anesthesia crew member would have to use that device to clear the patient's airway. Thus the shared physical event space became a medium of communication for the team. The prerequisite, of course, was the ability to understand the significance of patient events. The workspace itself is also a medium through which the teams coordinated. We often observed that team members, while not under instruction to perform specific tasks, scanned the workspace and perceived tasks needed to be carried out. In one case, for example, upon seeing an unopened package which would be used soon, a team member began to open the package and set up the device inside the package.

Activity monitoring. The interdependencies of tasks shared by a team mean that one member's tasks could sometimes only commence after the success of another member's tasks. (For example, surgeons can only begin certain procedures of resuscitation after the patient is anesthetized.) Thus monitoring the progress of an other member's tasks not only made it possible to compensate for a teammate's performance, but also gave lead information to prepare for the next step.

In many cases, the surgical crew did not announce their plans. However, the anesthesia crew inferred what needed to be done from the activities of the other crew. For example, during the review of the video tapes of a case,

one participant in that case revealed that the conversation between two surgical crew members provided cues of what the surgical crew would do next, even though the conversation was not directed at the anesthesia crew.

These strategies of task coordination, *without* the use of explicit verbal or gestural communications, enabled the resuscitation teams to perform smoothly in most situations.

Information Flow

One of the most interesting aspects of team coordination is the explicit, verbal communications regarding situational assessment and future plans, even though such communications were relatively rare (Figure 4). In the situations where such information flow was detected, we found most of the communications gave clear indications that the team was at a decision point. The team members voluntarily provided their views of the situation based on the decisions that the teams were facing at the time. For example in Figure 4, in one case when the patient was still not paralyzed 90 seconds (the usually sufficient duration) after the injection of drugs, several team members, without request, provided their assessment of the patient condition and of the reasons why the patient had not been paralyzed. In another case, while an anesthesia care provider was determining whether the patient was receiving oxygen, the surgeon provided his assessment of the situation unsolicited by saying that "the patient was stable."

The amount of verbal communications varied greatly among different teams. Some team leaders verbalized their plans clearly while other team leaders let the team members to infer their goals and intentions through actions.

Coordination Breakdowns

Considering the uncertainty and task difficulties involved in trauma patient resuscitation, the team coordination was adequate in the majority of the cases we analyzed. However, breakdowns in team coordination were observed in a number of crisis situations. We will report these breakdowns in the following three types of situations: (1) when there was pressure to seek alternative solutions, (2) when an unexpected, non-routine procedure was initiated, and (3) when there was a diffusion of responsibility.

Pressure to seek alternative solutions. In these type of situations, extreme difficulties or unexpected patient responses were encountered and prevented the implementation of routine procedures. When the patient condition was deteriorating rapidly, the team was under pressure to find an alternative solution and to act immediately. Figure 5 illustrates one such incident. In this case the patient had a gun shot wound to the lower abdomen. The patient's condition required immediate intubation (the passage of a tracheal breathing tube) to enable controlled ventilation, which required paralyzing the patient. The regular route to achieve this for the anesthesia crew was to wait for the surgical crew to gain venous access to the patient (phase A in Figure 5), as drugs to paralyze the patient were usually injected intravenously. However, difficulty in achieving this (due to previous use of veins for intravenous drug abuse) and rapidly declining patient conditions (unrecordable blood pressure, weak pulse, and combativeness due to agonal status) forced the anesthesia crew (with two members, ACP1 and ACP2) to examine alternatives.

During phase B (which represented a length of 20 seconds), the two anesthesia crew members implemented a line of action conflicting with each other's action. No attempt was made by either anesthesia crew member to communicate the problems or discuss action plans during this phase. The intentions and the objectives of each anesthesia crew member could only be inferred after their action plans were started.

Initiation of unexpected, non-routine procedures. This type of incident arose when unexpected non-routine and novel solutions were attempted. During phase C in Figure 5, for example, one of the anesthesia crew members decided to use a non-routine method (nasal intubation) of achieving airway access. This method required special materials that had not been anticipated in advance by the supporting members of the team. No announcement was made about the adoption of the non-routine method. As a result, the ability of the supporting members of the team to provide assistance was compromised. Coordination breakdowns in this type of incident were marked by the lack of anticipatory help from the team members, delays in preparing materials, and unnecessary pauses in the team leader's activities to obtain assistance.

Diffusion of responsibility. In critical circumstances during patient resuscitation, a diagnostic procedure or a treatment plan may have to be abandoned if the patient condition is too unstable. Such changes in plans occur during crises and under great time pressure. The team may have difficulties in adjusting itself from a diagnostic mode to action mode. Figure 6 shows one type of such scenario. During phase A, the anesthesia crew (labeled as ACP in Figure 6) concentrated on determining a critical task condition (whether or not the patient's lungs were being oxygenated), during which time the surgical crew (S) was assessing the patient condition and the nursing crew (N) was standing by, ready to provide assistance. After about 5 minutes the patient condition became critical (due to the lack of oxygen input), and the anesthesia crew decided to abort the process of obtaining further diagnostic cues. A sudden change of action (removal of the endo-tracheal tube or ET tube) was taken, without informing the rest of the team in advance during phase A. The inability of the rest of the team to anticipate this sudden change in plan prevented them adjusting their responsibilities accordingly, and resulted in the omission of a critical step (applying cricoid pressure to prevent regurgitation of stomach content into the lungs after the ET tube was removed).

Summary and discussions of Study 1

To summarize the strategies of team coordination, verbal communications can be viewed as one of many media that the team used to communicate. These types of media include, in addition to utterance and explicit gestures, (1) activities, (2) workspace, (3) events, (4) foci of attention. These media were possible because team members worked in closed physical workspaces. Although not sufficient in all occasions, they usually provide an efficient means for the team to coordinate.

The coordination breakdowns that our video analysis identified can be described in the following four forms: (1) conflicting plans, (2) inadequate support in crisis situations, (3) inadequate verbalization of problems, and (4) lack of task delegation. Their occurrence indicates gaps between what was needed and what the team had done in terms of team coordination.

The video recordings in our study show that team coordination was achieved in most situations with minimum explicit, verbal communications. When team coordination broke down, it often occurred in situations where there was a lack of explicit communication. In the following, we evaluate these findings against three previous studies done by Serfaty et al (1993),

Orasanu (1990), and Segal (1994).

In studies of team coordination patterns under stressful and unstressful situations, Serfaty et al (1993) found that high-performance teams were able to adapt their coordination strategies in stressful situations to reduce the cost of explicit communications. It appears that the teams in our study had adapted to the implicit coordination due to the high workload in many situations. Although no quantitative comparison was made between high stress and low stress situations, our observations show that in non-stressful situations, verbal communications contained considerable amount of non-essential information, some of which did not relate directly to the case involved. Such an adaptation could probably be better explained by the adaptation of workload management, as described by Sperandio (1971) in his analysis of communications between air traffic controllers and pilots.

Orasanu (1990) also contrasts team activities between high and low performance teams. Her major finding was that the content of communications was different between high and low performance teams. High performance teams communicated explicitly about problems and plans. However, in the cases that we recorded the small amount of verbal communications did not allow us to compare across different scenarios.

Segal's study (1994) of non-verbal communications had similar findings to ours. He found that visual monitoring of team members' activities was an important part of team coordination. Through the analysis of visual checking patterns, Segal provided quantitative data to support the notion that visible activity is an essential part of team work.

There are several implications of our findings for workplace design. Similar to what Segal (1994) argues, one has to beware of implicit communication channels, as they had important roles in team coordination in our studies. Practitioners utilized various non-verbal media for coordination: through activity monitoring and through shared event space. These media have important functional roles, including allowing team members to compensate for team mates and to schedule their own activities. The ability to monitor on-going activities and events also enables the team to have a coherent shared mental model (Cannon-Bowers et al, 1991; Orasanu, 1990), thus team member could provide needed information and support without an explicit request.

Our findings also provide guidance to studies of team activities in simulated environments. On the one hand, the current study highlights the importance of non-verbal communications and various types of medium used in communication. Stripping these methods of communication away in a laboratory study, for example, could dramatically change how a team coordinates and could impose extra workload on the team. Consequently, the problems in coordination observed in such a simulated setting may have a very limited validity in settings like emergency rooms. On the other hand, the three types of scenarios where coordination breakdowns were observed could lead investigators to focus on these scenarios and understand more about coordination breakdowns.

STUDY 2. UNCERTAINTY IN RESUSCITATION AND TEAM COMMUNICATION

A key characteristic of trauma patient resuscitation is the uncertainty involved: there are many unknowns about the patient and the incoming workload is unpredictable. An analysis was conducted to determine the impact of uncertainty on team performance.

Because of the availability of video recordings, it was possible to view these case segments repeatedly and to compare with the review comments provided retrospectively by the case participants. The analysis was carried out in order to answer the question of "What is uncertain to the team specifically related to this case segment?" A list of items uncertain to the team was generated for each case segment and the items from all cases were sorted.

A wide range of sources of uncertainty were identified. To illustrate how various types of uncertainty may arise during resuscitation, one case segment previously mentioned is described here:

> An unconscious patient with gun-shot wounds was brought in and the initial resuscitation started immediately. Efforts were made to place intravenous lines and a breathing tube for securing the patient's airway. Because it was uncertain whether the patient's C-spine (neck) was injured (e.g. during the fall after being shot), the patient's neck was immobilized to protect it from further injuries, even though such maneuver increased the difficulty of placing the tube. When an attempt was made to place the tube, the patient was found to have clenched jaws and the patient needed to be paralyzed first by intravenous administration of a muscle relaxant. However, partly due to the patient's drug abuse history, delays occurred in placing intravenous access. Because of this delay and the uncertainty of time before establishing intravenous access, the team considered whether to use an alternative method of placing the breathing tube without intravenous access, or simply to wait. Two minutes after muscle relaxant was injected, the patient still had clenched jaws. The team was uncertain why the desired muscle relaxation has not occurred within the usual time of under one minute.

Based on the analysis of all case segments, a list of ways in which uncertainties during resuscitation, either reported mentioned directly by subject matter expert reviewers or identified by analysts, was summarized here:

- The mechanism and the extent of injury.
- The patient's prior medical history.
- Working status of patient monitors. In comparison to many industrial settings, the monitors used during resuscitation frequently produce artifactual readings for various reasons. For example, several case segments involved displaced patient monitor probes.
- The effect of treatment. In the case segment described above, the patient was in shock with reduced blood circulation and was hemorrhaging. It was difficult to predict drug effects and identify the delays in onset under such circumstances.
- The availability of team members. There were several case segments in which the team was waiting for a member, uncertain when he or she would arrive. Possibilities for such uncertainty include that the members may have been with other patients. Concurrent patient admissions also made it difficult in several cases to predict the availability of team members, and consequently to decide whether to take measures to ameliorate potential adverse effects of reduced personnel resources.
- Task distribution among team members. Although each team member in the studied center has a nominally defined role, many case segments involved "negotiation" of who should do what. The frequent changes in team composition, the presence of personnel in training, and occasionally over-staffed resuscitation teams were three of the possible reasons.
- The intention of other team members.
- The resources and schedules of other parts of the institution (e.g. the availability of operating rooms and diagnostic devices).

- What has occurred during the time from injury to arrival in the trauma center and the status of the patient during field management and transport.

A total of 76 uncertain items were identified in 40 cases by examining verbal communications and subject matter expert reviews. They were categorized as:

- Patient Related Uncertainty (26%): Mechanism and extent of injury (8/76), reports by pre-hospital care providers (1/76), patient's prior medical history (5/76), effect of treatment in traumatized patient (6/76)
- Team/Organization Related Uncertainty (41%): resources and schedules (6/76), status/availability of team members (8/76), task distribution among team members (10/76), intention of other team members (7/76)

Interestingly, most of the uncertain items were related to team activities and the organization as whole as opposed to those related to patient conditions. This observation led to the possibility of improving performance by better intra-team communication (e.g. standardization, check-list) and by using data/video transmission of patient data (e.g. from the field).

Discussion of Study 2

Trauma patients, by definition, have traumatic changes in physiology and anatomy. These changes always cause uncertainty about the site and extent of patient injury. Part of the expertise of the people working in the domain is to deal with such uncertainty. However, when the activities during resuscitation were examined beyond the strict context of decision making, this study found that there were a number of sources of uncertainty that the clinicians had to deal with. In particular, uncertainties related to team activities and the organization as whole stood out.

In comparison with some of industrial settings where job specification and work process are well spelled out and team structure is clear to every team member, medical practices lack formal work structures. The interactions among the individuals caring for a patient are mostly informal in nature and are usually not codified. Within subgroups of professions (such as surgical care providers or anesthesia care providers) there exist hierarchical structures, but the care providers of a patient as a whole are not subject to formal rules that govern the exchange of information and materials. The lack of extensive standard operating procedures only compounds the informal characteristics of medical practices. Observed procedures changed from one leading surgeon to the next. Duplicate execution of tasks, unattended critical tasks, lack of communication of key information, conflicting plans, frustration of having to second-guess others' intention, etc, have been reported (e.g. Mackenzie et al, 1994; Donchin, et al, 1995, Xiao et al, 1995, Mackenzie, 1996).

It seems that lack of communication among the team members and among the personnel working in different departments contribute to much of the uncertainty identified. Such uncertainties could be reduced by, for example, explicitly informing other team members one's intention and plans. Improved communications on schedules and availability of resources (e.g. CT scanners) should also reduce uncertainties. Another source of uncertainties identified here was from patient monitors. Patient monitors are subject to many interfering factors, which often make monitors render faulty readings.

Little past literature examines uncertainty with the consideration of team and organization factors. A noteworthy recent attempt along this direction was made by Hutchins (1995). He postulated that, in order for a team to deal with varying task workload, there must be overlap among the team members' capabilities. As a consequence, there are remaining degrees of freedom in terms of task distributions among team members, which leads to uncertainties in who should do what when.

IMPLICATIONS AND RESEARCH NEEDS

The issue of temporal scale

Through our analysis of videotaped performance and our observation of collaboration in trauma patient care, it appears to be useful to view collaboration in three somewhat overlapping categories of collaborative efforts in terms of temporal scope:

- seconds-minutes
- hours-shift, and
- day-weeks

At different temporal scopes, different issues are at hand. For example, activity sequencing may be an important factor for collaboration over the span of seconds-minutes whereas development of team performance norms would be a concern when temporal scope is day-weeks. In the case of a trauma patient resuscitation team, members may be concerned with physical access to the patient and activities coordination over the span of seconds-minutes, but they may be concerned with assignment of roles and workload negotiation over the shift (hours-shift). Any concerns of procedures and staffing levels will be considered over the span of day-weeks.

Although they are highly related, emphasis of research about collaborative efforts over these three different temporal scopes can and should be different, as the fundamental issues of collaboration at these three temporal scopes are different in terms of training, design of information technology solutions, and research methodology. The studies reported above are mostly on the efforts in the temporal scope of seconds-minutes. Few would doubt that much can be learned by expanding the temporal scope to hours-shifts. Recent advances in information technology, for example, could be used to improve collaboration over longer time span.

Research needs

Human performance in regular context is almost always collaborative: individual performance is usually undefined when actual performance in any work settings is analyzed. In fact, performance should always be viewed as a result of collaboration among people in the context of tool using. In contrast, studies of human performance have not provided the research methodologies, framework, and descriptive languages for performance observed in actual work settings. Such contrast can be seen readily in our studies. Systematic and methodical analysis of video-based performance data is needed, and synthesizing and measuring instruments need to be established and tested.

The research community in the areas of human factors and cognitive engineering has started to face the challenge of studying performance in actual work environment. In medical domains, the increasing use of telecommunication and computation tools brought the issue of collaboration to the forefront. Previously isolated components in medical systems are more and more interconnected. Video-based studies on performance in real environment, as demonstrated above, can contribute much to the understanding of collaborative work.

ACKNOWLEDGMENT

Supported by funding from ONR N00014-91-J-1540,

AASERT, NASA NCC-2-921, ARL DAAL01-96-C0091, and Dynamics Research Corporation. The views and opinions of findings are the authors, not an official position or policy of the U.S. Department of Defense or NASA. LOTAS Group: Colin Mackenzie (Chair), W. Bernhard, C. Boehm, J. Blenko, H. Cline, D. Donnelly, R. Dutton, P. Hu, M. Jaberi, M. Kalish, K. Mitchell, and Y. Xiao.

REFERENCES

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., and Converse, S. A. (1991). Toward an integration of training theory and technique. *Human Factors*, **33**:281-292.

Hutchins, E. (1989). The technology of team navigation. In: Galeglier, J., Kraut, R. E., and Egido, C. (Eds.), *Intellectual Teamwork: Social and Technical Bases of Cooperative Work*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Kanki, B. G., Lozito, S., and Foushee, H. C. (1989). Communication indices of crew coordination. *Aviation, Space, and Environmental Medicine*, **60**:56-60.

Mackenzie, C. F., Craig, G. R., Parr, M. J., Horst, R., and the LOTAS Group. (1994). Video analysis of two emergency tracheal intubations identifies flawed decision making. *Anesthesiology*, **81**:911-919.

Mackenzie, C. F., Hu, P. F., Horst, R. L., and the LOTAS Group. (1995). An audio-video acquisition system for automated remote monitoring in the clinical environment. *Journal of Clinical Monitoring*, **11**:335-341.

Mackenzie, C. F., Martin, P., Xiao, Y., and the LOTAS Group. (1996). Video analysis of prolonged uncorrected esophageal intubation. *Anesthesiology*, **84**:1394-1503.

Morrissette, J. O., Hornseth, J. P., and Shellar, K. (1975). Team organization and monitoring performance. *Human Factors*, **17**:296-300.

Orasanu, J. (1990). *Shared mental models and crew decision making*. Tech. Rept. 46. Cognitive Science Laboratory, Princeton University, Princeton, NJ.

Orasanu, J., and Salas, E. (1993). Team decision making in complex environments. In: Klein, G. A., Orasanu, J., Calderwood, R., and Zsambok, C. E. (Eds.), *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex.

Rogalski, J., and Samurcay, R. (1993). Analysing communication in complex distributed decision-making. *Ergonomics*, **36**:1329-1343.

Segal, L. D. (1994). Actions speak louder than words: How pilots use nonverbal information for crew communications. In: *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, 21-25.

Serfaty, D., Entin, E. E., and Volpe, C. (1993). Adaptation to stress in team decision-making and coordination. In: *Proceedings of the Human Factors and Ergnomics Society 37th Annual meeting*, 1228-1232.

Sperandio, J. C. (1971). Variation of operator's strategies and regulating effects on workload. *Ergonomics*, **14**, 571-577.

Xiao, Y., Mackenzie, C. F., and the LOTAS Group. (1995). Decision making in dynamic environments: Fixation errors and their causes. In: *Proceedings of Human Factors and Ergonomics Society 39th Annual Meeting*, 469-473.

Xiao, Y., Hunter, W., Mackenzie, C., Jefferies, N., Horst R., and the LOTAS Group. (1996). Task complexity in emergency medical care and its implications for team coordination. *Human Factors*, **38**:636-645.

Zsambok, C. E., and Klein, G. A. (Eds.). (1997). *Naturalistic Decision Making*. Hillsdale, NJ: Lawrence Erlbaum Associates.

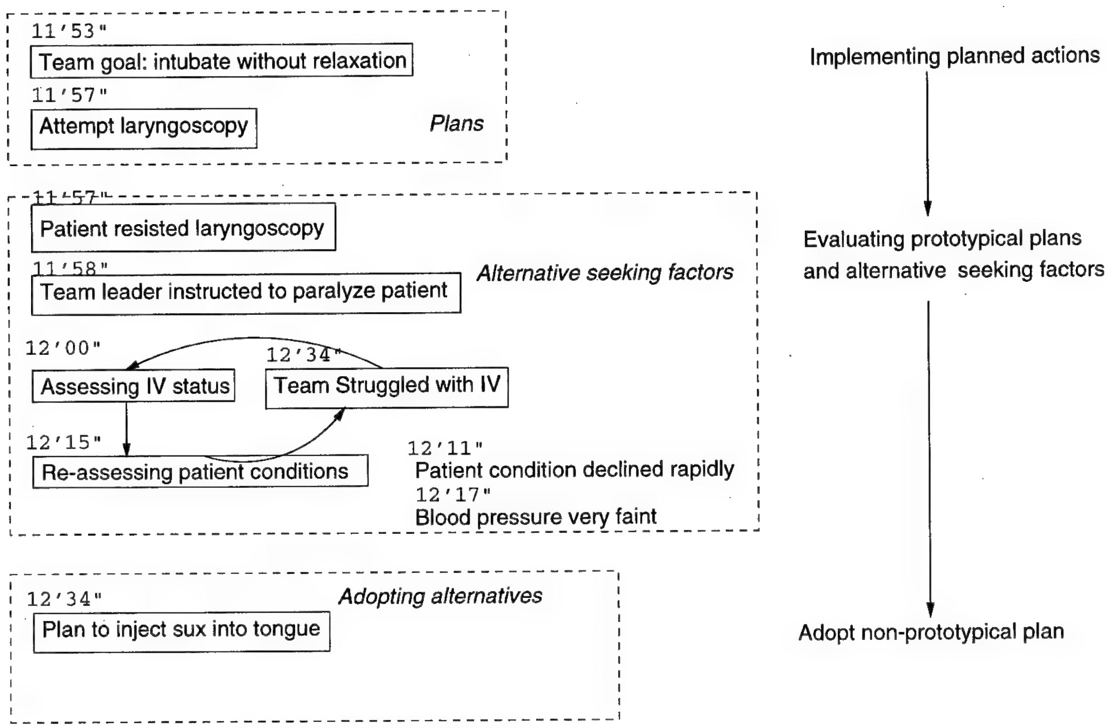


Figure 2. Sample event flow analysis. The time stamps are in min'sec". Abstraction of the events is represented on the righthand side.

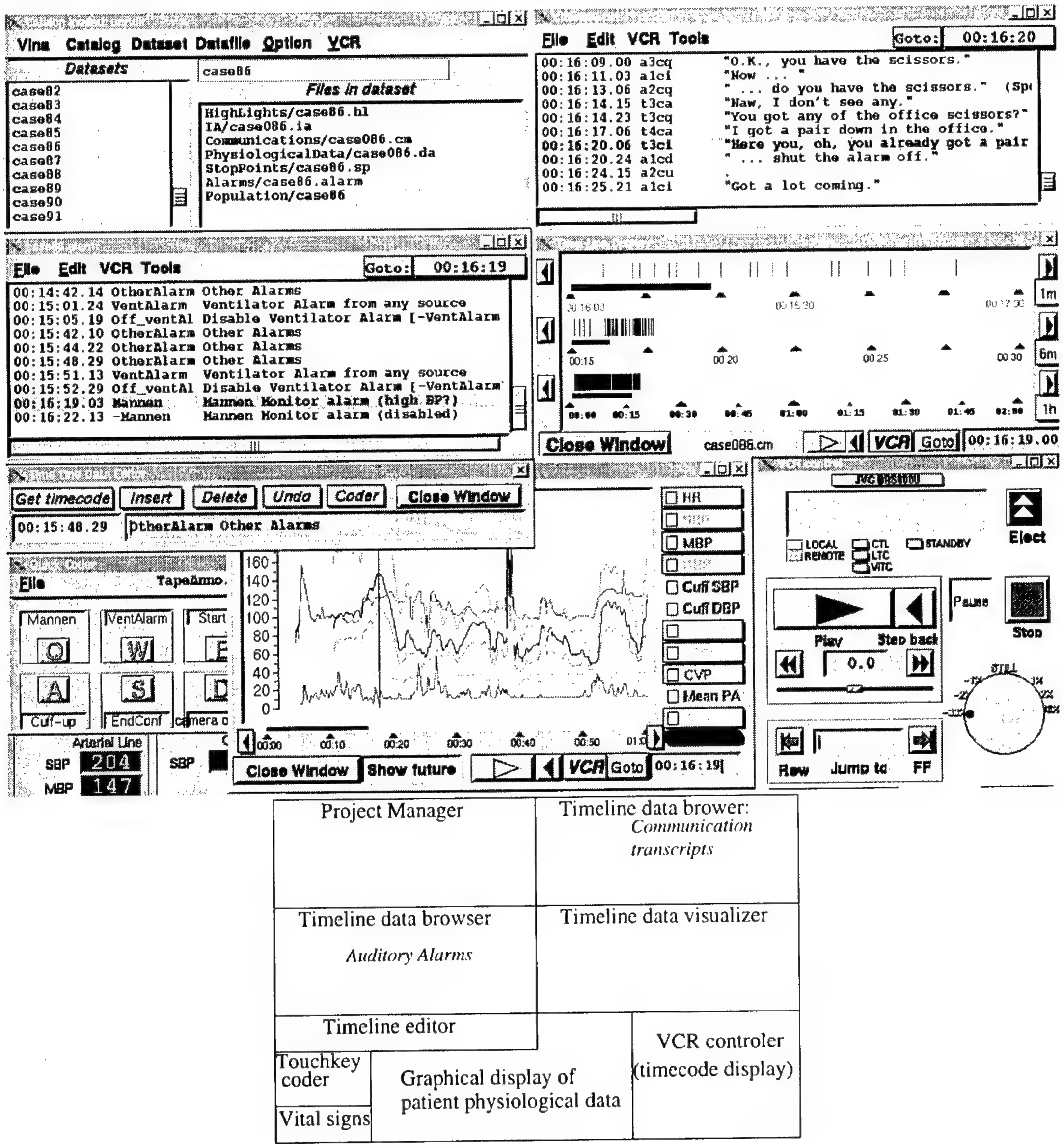


Figure 3. VINA: An environment for video analysis. Shown here is a screen dump (top) during the coding of auditory alarm events. The screen dump is explained in the bottom layout diagram. Current lines in landmark event window, the coding window, and the transcription window are automatically high-lighted to correspond the time code read directly from the VCR. The touch coder allows touch coding without stopping VCR. The graphical display of vital signs allows detection of abnormal trends and can position videotapes at interesting points with a click of a mouse button.

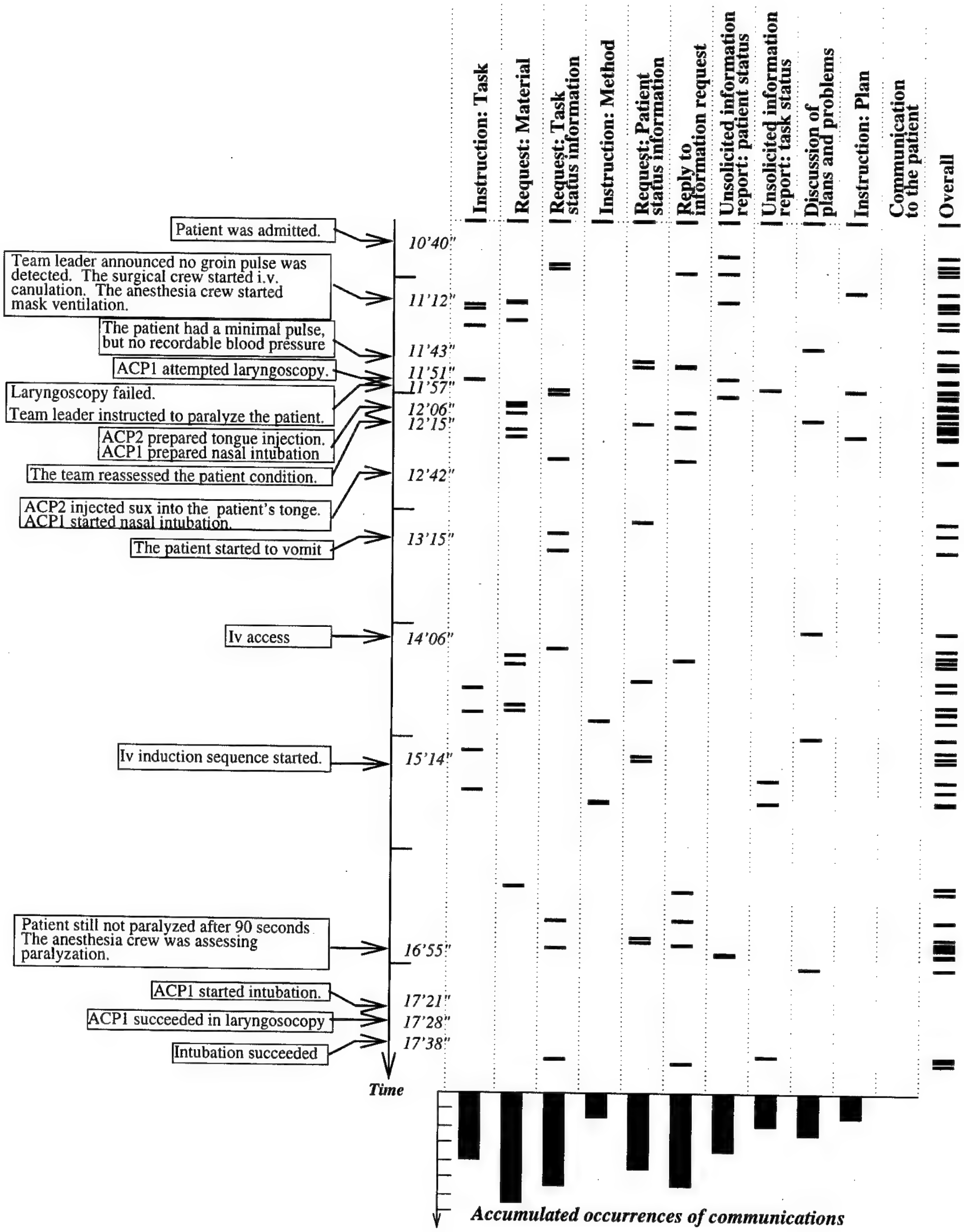


Figure 4. Sample analysis results of verbal communications. Each communication is shown by overall occurrence (right column) and categorized in the other 11 columns, using horizontal bars to depict their timings. The left column provides the time code, and key events are described on the vertical axis. The horizontal axis shows a histogram summarising the relative number of communications by category between patient admission time and successful intubation. ACP: anesthesia care provider, IV: intravenous.

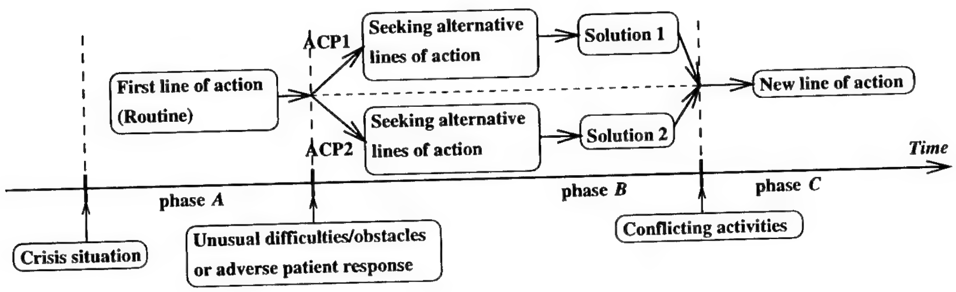


Figure 5. Coordination breakdowns when team encountering unexpected obstacle(s). Two anesthesia care providers are labelled as ACP1 and ACP2.

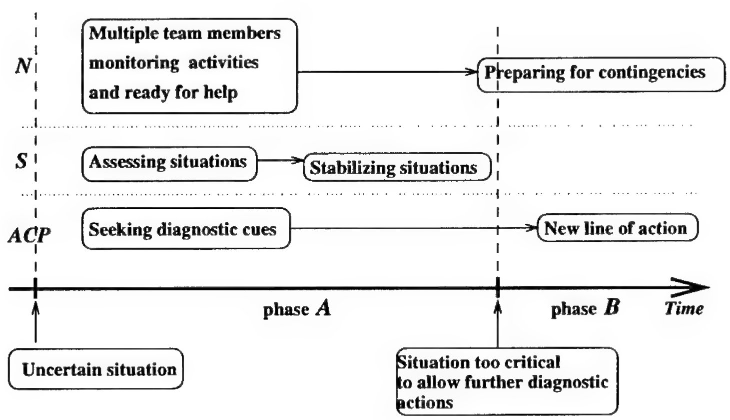


Figure 6. Coordination breakdowns when a sudden change of action occurred. N, S, and ACP represent three lines of activities of the nursing, surgical, and anesthesia crews, respectively.

Crew Concept WS-TORNADO - Navy (GE)

Jörg Schweingruber
Research Establishment for Applied Science (FGAN)
Research Institute for Electronics and Mathematics
Ergonomics and Information Systems
Neuenahrer Strasse 20
D - 53343 Wachtberg - Werthhoven / Germany
Tel.: +49.228.9435-491
Fax: +49.228.9435-508
e-mail: schweingruber@fgan.de

0 SUMMARY

The TORNADO aircraft weapon system that went into service with the German Navy in the middle of 1982 was based on the operational requirements and technical standards of the 70's. Changes in the operational goals forced requirements for improving, conserving and adjusting performance capabilities and made various equipments necessary being added to the current system. During the realisation of these modifications the ergonomic aspects of the TORNADO's man-machine interface, i.e., the cockpit, was largely neglected.

An analysis and evaluation of man-machine interactions in the cockpit was carried out including analysis of tasks, loads and demands on crew members and man-machine task allocation, depending on various typical navy missions and mission phases. In addition, extensive workload experiments were conducted with simulator and real flights representative for navy missions. Demands on subjects in simulator flights were analysed with questionnaires and the relative subjective workload measurement method SWORD, and in real flights with questionnaires and the absolute subjective workload measurement method ZEIS.

Experimental results indicated very high workload demands on crew members. Furthermore, extensive recommendations were established for reducing workload, so that mission requirements could be fulfilled. This resulted in a modified display concept with integrated video and FLIR-Sensor, modified warning displays and software for weapon selection as well as an integration of an additional radio and an instrument landing system. Beside these technical aspects social aspects, i.e., rank and level of training, have to be considered for optimizing crew coordination.

1 INTRODUCTION

The TORNADO aircraft weapon system had its first deployment in the German Navy in 1982. But the complete technical design based on the operational requirements and technical standards of the 70's or earlier (Schweingruber, 1995). A strict task separation was intended between front cockpit and rear cockpit. Changes in the operational goals forced requirements for improving, conserving, and adjusting performance capabilities and made various equipments necessary being added to the current system. These technical modifications led to changes in tasks and task allocations and the ergonomic aspects of the TORNADO's man-machine interface, i.e., the cockpit, were largely neglected (Schweingruber et al., 1995). This resulted in a temporary overload of crew members and a partly reduced system performance (Schweingruber et al., 1996).

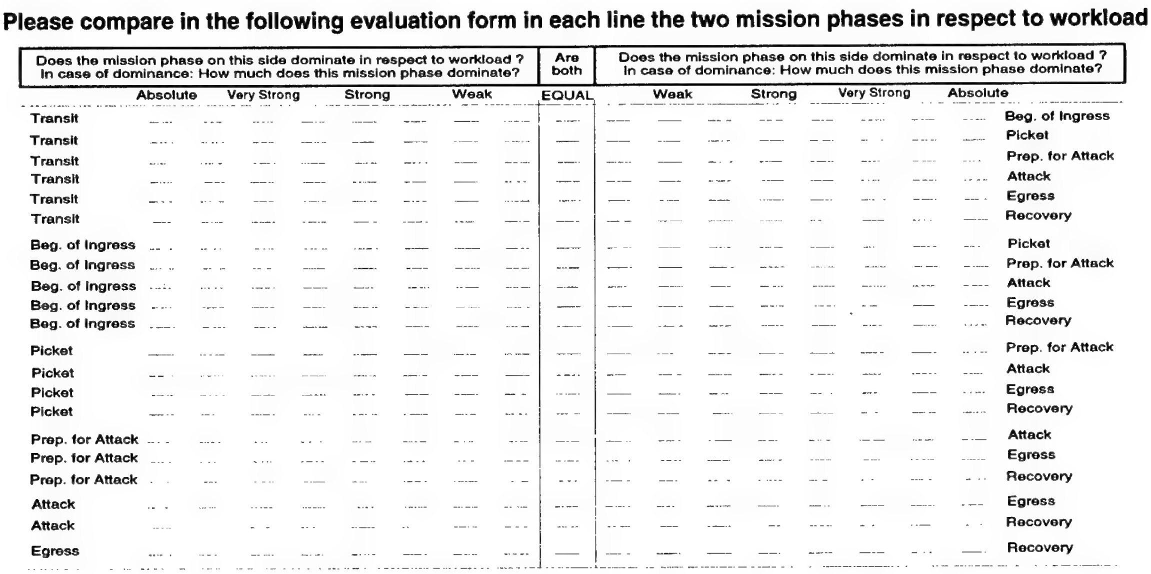
An analysis and evaluation of man-machine interactions in the cockpit was carried out, including analysis of tasks, input load, workload and demands on crew members and man-machine task allocation, depending on various typical navy missions and mission phases (Schweingruber, 1996). In addition, extensive workload experiments were conducted with simulator and real flights, representative for navy missions. The main goal was the reduction of input load and workload. Demands on subjects in simulator flight missions were measured with questionnaires and the relative subjective workload measurement method SWORD, and in real flight missions with questionnaires and the absolute subjective workload measurement method ZEIS (Hicks & Wierwille, 1979; Pfendler, 1982; Schick & Radtke, 1979; Pfendler & Schweingruber, 1996).

2 SIMULATOR FLIGHT MISSIONS (Experiment I)

Two representative navy missions were performed in a TORNADO simulator to determine the workload profile, the mission phases with workload peaks, and the input load factors. In one type of mission, missiles (Kormoran) were employed by 11 crews, and in the other one bombs by 14 crews.

For this reason the Subjective WORkload Dominance Technique (SWORD) (Saaty, 1980; Vidulich, 1989) was used to rate workload of the crews in 7 defined mission phases (Figure 1) (Siegel, 1976). Furthermore questionaires and interviews were carried out with the navy instructors referring to the individual mission success and with the crew members referring to their input load and workload.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.


Please compare in the following evaluation form in each line the two mission phases in respect to workload

Does the mission phase on this side dominate in respect to workload ? In case of dominance: How much does this mission phase dominate?	Are both	Does the mission phase on this side dominate in respect to workload ? In case of dominance: How much does this mission phase dominate?
Absolute / Very Strong / Strong / Weak	EQUAL	Weak / Strong / Very Strong / Absolute
Transit		Beg. of Ingress
Transit		Picket
Transit		Prep. for Attack
Transit		Attack
Transit		Egress
Transit		Recovery
Beg. of Ingress		Picket
Beg. of Ingress		Prep. for Attack
Beg. of Ingress		Attack
Beg. of Ingress		Egress
Beg. of Ingress		Recovery
Picket		Prep. for Attack
Picket		Attack
Picket		Egress
Picket		Recovery
Prep. for Attack		Attack
Prep. for Attack		Egress
Prep. for Attack		Recovery
Attack		Egress
Attack		Recovery
Egress		Recovery

Figure 1: SWORD Rating Scale

In contrast to most other techniques, which are based on absolute comparisons, the SWORD method is based on paired comparisons (Vidulich et al., 1991). This means that each task (mission phase) has to be compared individually to all other tasks (other mission phases) retrospectively after the complete mission (Pfendler et al., 1995; Pfendler et al., 1997; Schweingruber & Pfendler, 1997).

3 REAL FLIGHT MISSIONS (Experiment II)

The real flight missions were carried out on a navy air force base, with TORNADO navy aircrafts under combat training conditions, in a wide range of representative typical navy missions, with 32 pilots and 30 weapon system officers, to work out the maximum workload of typical navy missions, typical input load factors and the load effects on system performance.

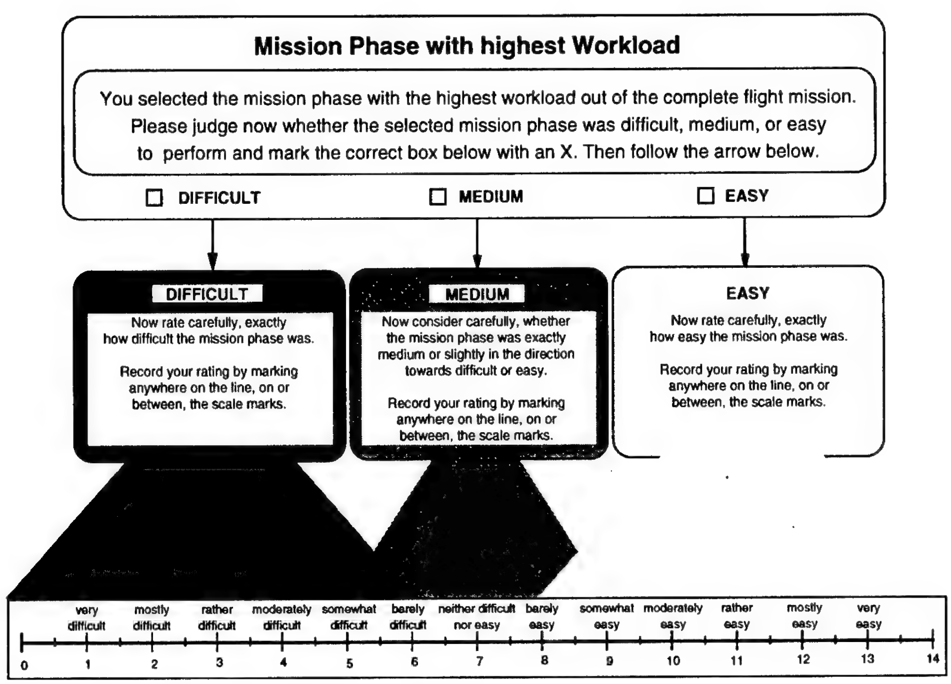


Figure 2: ZEIS Rating Scale

For this reason the subjective workload measurement method ZEIS, a sequential judgement scale, was applied (Käppler et al., 1988). Crews had to rate workload of the mission phase with the highest workload during each complete indiviual navy mission. Furthermore questionaires and interviews were carried out with the crew members reffering to input load and workload during the missions.

In contrast to the SWORD method, ZEIS is a sequential judgement scale based on absolute ratings (Pitrella, 1989). This means that each task requires two judgements in sequence, first a coarse judgement, and then a second finer one (Figure 2). The first judgement is made according to the three basic categories 'difficult', 'medium' or 'easy'. With specific instructions the second judgement is made on a smaller section of the full continous scale.

4 RESULTS

4.1 Workload Profile (Results from SWORD / Experiment I)

The workload profile from SWORD (Figure 3) for both navy missions (Kormoran and bomb) and for both crew members (pilot and weapon system officer) indicates a similar increase of workload from mission phase 1 (transit) to mission phase 5 (attack) with a maximum workload during mission phase 5 and a workload decrement from mission phase 5 (attack) to mission phase 7 (recovery). The differences of the profiles between pilots and weapon system officers in mission phase 3 (picket) and mission phase 4 (preparation for attack) are depending on the different tasks of the crew members in this mission phases because of the different weapons.

4.2 Maximum Workload (Results from ZEIS / Experiment II)

The maximum level of the ZEIS workload profile (Lilliefors, 1967; Dallal & Wilkinson, 1986; Mason & Bell, 1986; Sachs, 1992) indicates a higher workload for the pilots than for the weapon system officers (Figure 4). The difference is, according to the arithmetic means, exactly 1 Step of 15 Steps. Furthermore the workload profile of the weapon system officers is more wide and

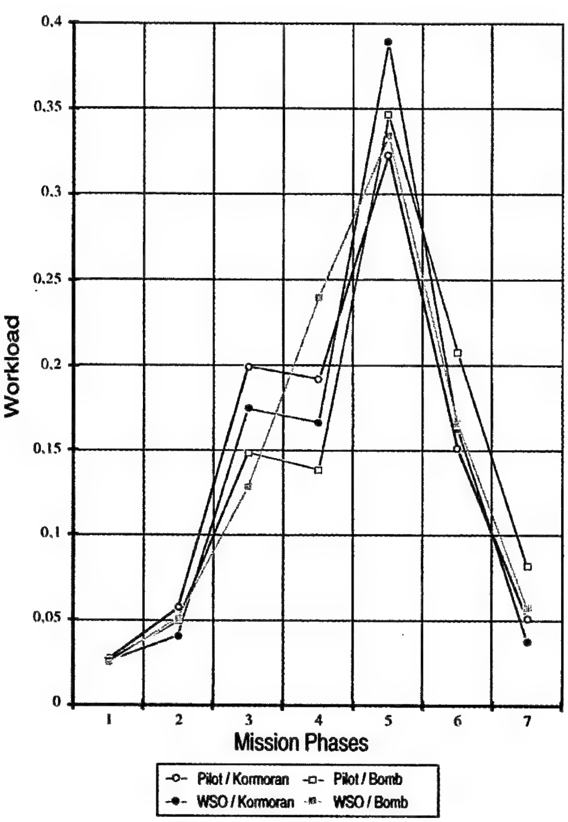


Figure 3: Workload Profile (SWORD / Experiment I)

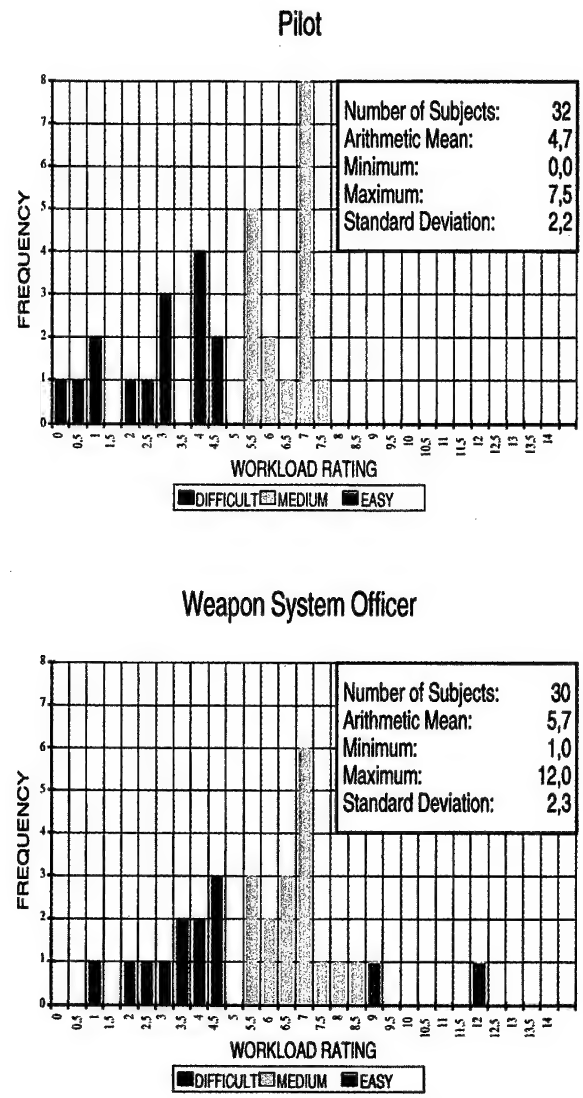


Figure 4: Maximum Workload (ZEIS / Experiment II)

more in the lower workload region than the workload profile of the pilots. This means that there is a higher workload for the pilots than for the weapon system officers in the mission phase with the highest workload.

4.3 Input Load, Workload and Recommendations for Workload Reduction (Results from Questionaires and Interviews)

Refering to input load and workload factors the questionaires and interviews showed problems with intecommunication depending on training level, information deficits of pilots regarding to situation status which should be transmitted from the rear cockpit to the front cockpit and increasing workload while crews trying to flexibilize the task allocation.

To reduce input load and workload, a more flexible task allocation with equipment support for the crew members and more simplified system input procedures are recommended.

The technical modifications to reduce workload should include a multi-function head-down display in the front cockpit and the rear cockpit, a head-down display in the front cockpit with integrated video sensor and forward-looking infrared (FLIR) sensor informations, identical warning display units in the front cockpit and the rear cockpit, a simplification and standardization of the weapon launch procedures for all weapons, hand grips combined with buttons for countermeasures and radio in the rear cockpit, an instrument landing system and a second radio.

5 CONCLUSIONS

Experimental results indicated very high workload demands on crew members. Extensive recommendations were established for reducing workload to ensure required performance levels. This resulted in a changed display concept with integrated video and FLIR-Sensor, modified warning displays and software for weapon selections as well as an integration of an additional radio and an instrument landing system. Beside these technical aspects social aspects, i.e., rank and level of training, have to be considered for optimizing the crew coordination. All this recommendations for reducing workload are only effective, if they are integrated in a comprehensive ergonomic cockpit design and crew coordination concept.

6 REFERENCES

Dallal, G. E., Wilkinson L. (1986): *An analytic approximation to the distribution of Lilliefors's test statistic of normality.* The American Statistican 40, 294 - 295.

Hicks, Th. G., Wierwille, W. W. (1979): *Comparison of Five Mental Workload Assessment Procedures in a Moving-Base Driving Simulator.* Human Factors, 1979, 21(2), S. 129-143.

Käppler, W. D., Pitrella, F. D., Godthelp, H. (1988): *Psychometric and Performance Measurement of Light Weight Truck Handling Qualities.* Bericht Nr. 77. Forschungsinstitut für Anthropotechnik. Wachtberg.

Lilliefors, H. W. (1967): *On the Kolmogorov-Smirnov test for normality with mean and variance unknown.* J. Amer. Statist Assoc., 399 - 402.

Mason, A. L., Bell, C. B. (1986): *New Liliefors and Srinivasan tables with applications.* Communication in Statistics - Simulation and Computation 15, 451 - 477.

Pfendler, C. (1982): Bewertung der Brauchbarkeit von Methoden zur Messung der mentalen Beanspruchung bei Kfz-Lenkaufgaben. Z. Arb. wiss. 36 (8 NF) 1982/3, S. 170-174.

Pfendler, C., Pitrella, F. D., Wiegand, D. (1995): *Messung der Beanspruchung bei der Systembewertung.* Bericht Nr. 115. Forschungsinstitut für Anthropotechnik. Wachtberg.

Pfendler, C. & Schweingruber, J. (1996): *Übersicht über rechnergestützte Methoden zur Beanspruchungsermittlung.* Forschungsbericht Nr. 467. Forschungsinstitut für Funk und Mathematik. Wachtberg.

Pfendler, C., Schweingruber, J. & Huland, W.-D. (1997): *Beanspruchungsermittlung mit der Subjective Workload Technique (SWORD) und ihrer rechnergestützten Version.* Forschungsbericht Nr. 477. Forschungsinstitut für Funk und Mathematik. Wachtberg.

Pitrella, F. D. (1989): *Validation of the 15 point German language Sequential Judgment Scale.* unveröffentlicht. Forschungsinstitut für Anthropotechnik. Wachtberg.

Saaty, T. L. (1980): *The Analytic Hierarchy Process.* NewYork: McGraw-Hill.

Sachs, L. (1992): *Angewandte Statistik.* Berlin: Springer.

Schick, F. V., Radtke, H. (1979). *Untersuchung der Pulsfrequenzvariabilität als Schätzgröße der Pilotenbeanspruchung bei anthropotechnischen Experimenten.* DFVLR - FB 79-33.

Schweingruber, J. (1995): *Besatzungskonzept WS-TORNADO - Marine.* Zwischenbericht. Forschungsinstitut für Anthropotechnik. Wachtberg.

Schweingruber, J.; Eßer, O.; Käppler, W. D. & Gärtner, K.-P. (1995): *Besatzungskonzept WS-TORNADO - Marine.* Abschlußbericht 1995. Forschungsinstitut für Anthropotechnik. Wachtberg.

Schweingruber, J.; Eßer, O.; Käppler, W. D. & Gärtner, K.-P. (1996): *Ermittlung der Belastung und Beanspruchung von Besatzungsmitgliedern im Waffensystem TORNADO - Marine.* Forschungsbericht Nr. 458. Forschungsinstitut für Funk und Mathematik. Wachtberg.

Schweingruber, J. (1996): Belastung und Beanspruchung von Flugzeugführern im Zwei-Mann-Cockpit eines militärischen Flugzeugmusters. In: K.-P. Gärtner (Hrsg.). *Interaktive Informationssysteme in Fahrzeugen und Leitstellen.* DGLR-Bericht 96-02. Deutsche Gesellschaft für Luft- und Raumfahrt e.V. Bonn.

Schweingruber, J. & Pfendler C. (1997): Messung der Beanspruchung mit der rechnergestützten Version von SWORD. In: K.-P. Gärtner (Hrsg.). *Menschliche Zuverlässigkeit, Beanspruchung und benutzerzentrierte Automatisierung.* DGLR-Bericht 97-02. Deutsche Gesellschaft für Luft- und Raumfahrt e.V. Bonn.

Siegel, S. (1976): *Nicht parametrische statistische Methoden*. Frankfurt am Main: Fachbuchhandlung für Psychologie, Verlagsabteilung.

Vidulich, M.A. (1989): *The Use of Judgement Matrices in Subjective Workload Assessment: The Subjective Workload Dominance (SWORD) Technique*. Proceedings of the Human Factors Society 33rd Annual Meeting. Santa Monica, CA: Human Factors Society, pp. 1406-1410.

Vidulich, M.A., Ward, G.F. and Schueren, J. (1991): *Using the Subjective Workload Dominance (SWORD) Technique for Projective Workload Assessment*. Human Factors, 33(6), pp. 677-691.

TEAM COMPATIBILITY AS A PREDICTOR OF TEAM PERFORMANCE: PICKING THE BEST TEAM

Gary G. Kay
Director, Neuropsychology Division
Georgetown University Medical Center
3800 Reservoir Road, NW
Washington, DC 20007
USA

D. L. Dolgin
NAWC-AD, Crew Systems
NAS Patuxent River, MD USA

1. SUMMARY

Collaborative crew systems are likely to be influenced by the interpersonal relationships that exist between crew members. Although, individual members of crews are often highly screened and selected for their cognitive capacity, intelligence, conscientiousness, and emotional stability, little attention is generally paid to the interpersonal needs of team members or to the overall compatibility of the team. Schutz (1966) postulated that a group with higher compatibility will have higher goal achievement than a group with lower compatibility. The present study investigated the utility of a self-report personality test (the FIRO-B) as a measure of team compatibility. The teams participating in the study were 3-man groups of U.S. Air Force AWACS Weapon Directors engaged in high fidelity simulation of combat conditions. Results of the study generally did not support the hypothesis that high compatibility, as measured by the FIRO-B, would be associated with better simulated AWACS performance. Nevertheless, crew compatibility appears to have been a factor in team performance for several of the teams. The article stresses the need for developing effective measures of team compatibility.

2. INTRODUCTION

This presentation will address two questions: (1) How does compatibility affect team performance? and (2); Can results from an individually administered personality test designed to measure interpersonal needs be used to evaluate the extent to which teams are compatible? I will present personality test results (from the FIRO-B test) for twelve 3-man weapons director teams who participated in an AWACS simulation study in 1989.

Schutz (the author of the FIRO-B test)[1] stated the *Postulate of Compatibility* as follows: If the compatibility of one group, *h*, is greater than that of another group, *m*, then the goal achievement of *h* will exceed that of *m*. The present study was designed to test this postulate.

If the postulate of compatibility is correct we need to focus not only on selecting well qualified individuals, but we also need to select individuals who can be compatible within a group. Compatibility refers to the "the relations between two or more persons that leads to mutual satisfaction of interpersonal needs and harmonious existence."

There is increasing recognition of the role of compatibility in a variety of operational environments. Maximizing interpersonal effectiveness is a goal of CRM. In some operational environments, such as in Special Forces Teams, and for teams functioning in isolation and confinement

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

(e.g., long duration space flight (MIR, Antarctic Winter-Over, and Space Station) crew compatibility is particularly critical. In these environments we recognize the critical role that interpersonal factors play in operational performance. Investigators are looking closely at the interpersonal issues in these environments.

2.1 Background of the FIRO-B

Schutz began the work that led to the development of the FIRO-B while at the Naval Research Laboratory in Washington, DC in 1952. He had been assigned the task of understanding and improving the performance of the Combat Information Center (CIC) of Navy ships. Schutz's basic idea is that "every person orients himself [or herself] in characteristic ways toward other people, and that knowledge of these orientations allows for considerable understanding of individual behavior and the interaction of people." This idea is expressed as the Fundamental Interpersonal Relations Orientation (FIRO) and is measured by a test known as the FIRO-B.

The test generates 6 scores (0-9) indicating expressed and wanted needs in the areas of Inclusion, Control, and Affection. Scores from 0 to 1 are considered extremely low and scores from 8 to 9 are considered extremely high scores. The current study limited the analysis to the inclusion and control areas.

Inclusion: Refers to one's general social orientation. A low expressed score means that the person is uncomfortable around people and will tend to move away from them. A high expressed score suggests that the person is comfortable in social settings and will tend to move toward people. A low wanted inclusion means that the person is selective about with whom association takes place. A high wanted score means that the person has a strong need to belong and be accepted.

Control: Pertains to leadership behavior. A low expressed score means that the person avoids making decisions and taking on responsibility; a high expressed score indicates that the person can and does take on the responsibilities involved in a leadership role. A low wanted score suggests that the person does not want to be controlled by others. A high wanted score reflects abdication of responsibility and a disposition toward accepting control from others.

2.2 FIRO-B Measures of Compatibility

2.2.1 Reciprocal Compatibility

The degree to which the expressed behavior of one person equals the wanted behavior of the other person. - how well two people satisfy each other's needs.

$$rK_{ij} = [e_i - w_j] + [e_j - w_i]$$

Scores can range from 0-18.

2,2,2 Originator Compatibility

The degree to which individuals desire to initiate or be the recipient of interpersonal behaviors. In the control area: a preference for dominating and controlling others and strongly resisting their influence (originate only) as opposed to always being influenced and never being influential (receive only).

$$oK_{ij} = (e_i - w_i) + (e_j - w_j)$$

2.2.3 Interchange Compatibility

Extent to which team/group expresses inclusion, control or affection. Relates to the preferred amount of interchange.

$$xK_{ij} = [(e_i + w_i) - (e_j + w_j)]$$

3. STUDY DESIGN:

The study was conducted at the US Air Force, Armstrong Laboratory, Brooks Air Force Base, Texas. The laboratory was at that time conducting

studies investigating the impact of antihistamines on the cognitive and psychomotor skills of Airborne Warning and Control System (AWACS) Weapons Directors.[2] This work was being conducted with a high fidelity AWACS simulator. The simulation environment provided weapons directors with high and low workload combat scenarios. Eleven of the twelve 3-man Weapons Director teams who participated in the antihistamine study volunteered to complete the FIRO-B along with other questionnaires and personality instruments that were part of the primary study.

On the first day of the study, *Day 1*, subjects were trained on cognitive and psychomotor tests and introduced to the simulator. *Day 2* served as the baseline day. On *Day 3* teams received either terfenadine, placebo, or diphenhydramine. On *Day 4* the subjects continued taking the study medication. Weapons Directors completed both high and low workload scenarios.

3.1 AWACS Measures:

(1) Composite simulator score under high and low workload conditions

(2) Number of Penetrations

Data were used from Day 2 (baseline) and Day 4 (2nd antihistamine day). The data from Day 4 were considered acceptable for use in the study because of the lack of demonstrated drug effect on Day 4. The AWACS performance scores were ranked.

3.2 FIRO-B Measures:

Team data for each 3-man group of Weapons Directors was obtained by summing the compatibility scores for each of the three dyads formed by the team:

Subject A x Subject B
Subject A x Subject C
Subject B x Subject C

Team compatibility scores were then ranked. Lower scores indicate higher levels of compatibility.

Spearman rank-order correlations were performed to assess the relationship between compatibility scores and AWACS performance scores.

4. RESULTS:

Two significant correlations were obtained. Both correlations were in the opposite direction from that predicted.

Lower interchange compatibility was associated with better performance on the composite AWACS simulation score measure ($r = .60$; $p = .025$). Also, better performance under the low workload condition on Day 2 was associated with lower levels of inclusion compatibility ($r = .57$; $p = .033$).

Substantial, though non-significant correlations were obtained for other measures, in the predicted direction. The total number of penetrations decreased with better interchange compatibility ($r = .39$; $p = .12$) and control compatibility was associated with better AWACS performance on the composite measure under the high workload simulation condition on Day 4 ($r = .41$; $p = .11$).

Analysis of individual teams yielded further information regarding the extent to which the FIRO-B predicted team performance. Results for four of the teams appearing at the end of this paper (Annexes A-D) demonstrate interesting relationships between FIRO-B scores, compatibility, and simulation performance.

5. CONCLUSIONS:

Results from the present study are not particularly convincing with respect to the validity of the FIRO-B compatibility scores as predictors of operational performance. Specifically, FIRO-B compatibility measures did not appear to strongly predict performance in the AWACS simulator. In

fact, lower Interchange Compatibility was significantly associated with higher scores on the AWACS composite simulation measure. In spite of this paradoxical finding, this study should draw our attention to the need for further development of instruments for assessing the interpersonal characteristics of teams or crews that can impact on mission critical operations.

The FIRO-B has received considerable criticism regarding whether the test is a valid measure of interpersonal behavior or intrapersonal characteristics. Hurley[4] reported low levels of association between FIRO-B scores and peer-ratings on Lorr and McNair's (1963)[5] Interpersonal Behavior Inventory (IBI) for undergraduates participating in ten weeks (50 hours) of interpersonal skills groups. The FIRO-B Control dimension was found to have the strongest association with relevant scales on the IBI. Salminen (1991) reported disappointing results for the FIRO-B in an examination of the test's convergent and discriminant (i.e., construct) validity.[6] Only three of the six intercorrelations on the validity diagonal were statistically significant and only 80% of the discriminant validity comparisons met criteria.

In contrast, Fisher and colleagues[7], using factors obtained from a reformulation of Schutz' ideas on team compatibility found that "Group-Warmth", a derivative of FIRO-B Inclusion and Affection scales, was significantly related to the commercial effectiveness of teams.

Schutz has reexamined the FIRO-B, and made substantial changes in the instrument resulting in the development of Element B.[8] In Element B the Affection scales are replaced by a measure of Openness. Instead of Expressed and Wanted, Schutz now has the respondent describe Expressed, Received, Perceived and Wanted behavior. According to Schutz, when given in conjunction with other instruments, he claims that his new instruments have been used for "improving self-awareness, teamwork, morale, and productivity in such organizations as Proctor & Gamble, AT&T, NASA, Amdahl Corporation, the Swedish Army and about 100 companies in Japan (p.915)[8].

Selection of individuals tends to focus on screening-out psychopathology and screening-in intelligence and cognitive processing capacity. Very little attention is paid to selecting individuals based on their non-pathological interpersonal characteristics. The "best person for the job" may not necessarily be the "best person for the team."

The question remains open: Can self-report personality measures be used to form groups of individuals who are likely to be maximally compatible. Field studies and group/team simulation activities provide a direct means of observing teams engage in task-related activities. These studies also allow for the measurement of the team's productivity. However, the field-study approach to team assessment has considerable costs and doesn't easily permit the evaluation of all possibly permutations of team member groupings. A validated paper-and-pencil measure of interpersonal compatibility would be far less costly and time consuming. Also, a paper-and-pencil (self-administered) test would allow for the modeling of all possible groupings of potential team members.

There are measures of interpersonal functioning other than the FIRO-B. Buhrmester and colleagues[3] developed a questionnaire to assess five dimensions of interpersonal competence: initiating relationships, self-disclosure, asserting displeasure with others' actions, providing emotional support and managing interpersonal conflicts. Moderate levels of agreement were found between ratings of competence by subjects and their roommate ratings. NASA has also been developing measures of interpersonal characteristics that have been predictive of behavior in isolated Antarctic environments (Woods, J., personal communication, 1998).

Clearly, we need to match our advances in the selection of individuals with advances in methods for selecting and forming effective teams. It is likely that the instruments that work with one population or for one type of activity will not be effective elsewhere. As our measurement of interpersonal characteristics improves we will need to better understand the role of situational variables and the interaction of situational variables on the characteristics of teams.

6. ACKNOWLEDGMENTS

The authors thank Dr. Samuel Schiflett and Dr. Douglas Eddy for collaborating in this study, providing the AWACS data, and for collecting the FIRO-B data on the eleven AWACS teams. The authors also appreciate the material provided by Dr. Joanna Wood, Krug Life Sciences, Inc., and Dr. Michael McNeese, US Air Force.

7. REFERENCES:

1. Schutz, W. "The Interpersonal Underworld: FIRO A Three-Dimensional Theory of Interpersonal Behavior", Palo Alto, California, Science and Behavior Books, Inc., 1966 (ISBN: 0-8314-0011-0).

2. Nesthus, T.E., Schiflett, S.G., Eddy, D.R., & Whitmore, J.N. "Comparative Effects of Antihistamines on Aircrew Performance of Simple and Complex Tasks Under Sustained Operations", Armstrong Laboratory, US Air Force Systems Command, AL-TR-1991-0104, December, 1991.

3. Buhrmester, D., Furman, W., Wittenberg M.T., & Reis, H.T. "Five Domains of Interpersonal Competence in Peer Relationships", J. Pers. Soc. Psychol., *55*, December 1988, pp 991-1008.

4. Hurley, J.R. "Does FIRO-B Relate Better to Interpersonal or Intrapersonal Behavior?" J Clin Psychol, *46*, July 1990, pp 454-460.

5. Lorr, M., & McNair, D.M. "An Interpersonal Behavioral Circle", J Abnormal and Social Psychol, *67*, 1963, pp 68-75.

6. Salminen, S. "Convergent and discriminant validity of FIRO-B Questionnaire", Psychol Reports, *69*, 1991, pp 787-790.

7. Fisher, S.G., Macrosson, W.D.K., & Walker, C.A. "FIRO-B: The Power of Love and the Love of Power", Psychol Reports, *76*, 1995, pp 195-206.

8. Schutz, W. "Beyond FIRO-B-Three New Theory-Derived Measures-Element B: Behavior, Element F: Feelings, Element S: Self", Psychol Reports, *70*, 1992, pp 915-937.

ANNEX A.

Example of Highly Compatible Group (#6); rankings on each type of compatibility:

Reciprocal 3/8	Origin 1/6	Interchange 3/7
Control 1/5	Control 1/5	Control 2/6
Inclusion 3/4	Inclusion 2/6	Inclusion 5/8

Performance:

Composite Simulator Score 2/9
Penetrations 1/7

Comment: High compatibility in areas of Control and Origination Compatibility. Relatively young group with below average hours.

Members of group:

	Inclusion Expressed	Wanted	Control Expressed	Wanted
Subject A	4	8	3	2

I= Concerned about rejection, need to belong is intense, Highly sensitive to being left out, perceived by others as reasonably sociable. C= Not only avoids making decisions and taking responsibility, but is most conformable when others don't not attempt to control them. They don't tell others what to do. New, untried and untested areas make them anxious. Good potential for leadership, but takes responsibility at his own speed. Can't be rushed ("rebel" type).

	Inclusion Expressed	Wanted	Control Expressed	Wanted
Subject B	5	0	0	2

I=No need for constant socialization,. Have many acquaintances but very few persons with whom they care to spend any great time. Good social skills. C= Same as above (rebel).

	Inclusion Expressed	Wanted	Control Expressed	Wanted
Subject C	1	0	0	1

I=Most comfortable when can move away from people or when people in general stay away from him. Highly selective in associations. Uncomfortable around most people and avoids others when possible. Concerned about rejection. C= Same as above (rebel) "tend to associate mostly with other rebels".

ANNEX B.

Example of Moderately Low Compatibility Group (#2); rankings on each type of compatibility:

Reciprocal 6/8	Origin 5/6	Interchange 2/7
Control 4/5	Control 5/5	Control 4/6
Inclusion 3/4	Inclusion 1/6	Inclusion 4/8

Performance:

Composite Simulator Score 8/9
Penetrations 6/7

Comment: Low compatibility in areas of Reciprocal and Origination. Incompatibilities possibly in area of control. Also, team had lowest number of simulation hours and lowest number of years. However, overall number of hours higher than Team #6 (described in Figure 1).

Members of group:

	Inclusion Expressed	Inclusion Wanted	Control Expressed	Control Wanted
Subject A	2	1	1	1

I= Most comfortable when he can move away from people or when people stay away from him. Highly selective and uncomfortable around most people. Reject others before being rejected themselves. C= "Rebel" type.

	Inclusion Expressed	Inclusion Wanted	Control Expressed	Control Wanted
Subject B	7	6	9	0

I=Compulsively driven toward people. Difficult to be alone for extended periods of time. Needs to be accepted or belong, fear of rejection. Seeks out people and socializes with ease. C=Self concept is one of confidence and adequacy - walk into areas of responsibility where most angels fear to tread - Intense need for recognition - compulsively driven to do well. Compulsively take on large amounts of responsibility hoping they will earn recognition. Attracted to others who give them the recognition they desire, and to those who do not make decisions for them or attempt to control them.

	Inclusion Expressed	Inclusion Wanted	Control Expressed	Control Wanted
Subject C	6	9	1	1

I=Compulsively driven toward people. Difficult to be alone for extended periods of time. Needs to be accepted or belong, fear of rejection. Seeks out people and socializes with ease. C= "Rebel" type.

ANNEX C.

Example of team that appeared compatible on FIRO-B but performed poorly on AWACS (Team #3):

Reciprocal 2/8	Origin 3/6	Interchange 1/7
Control 2/5	Control 3/5	Control 1/6
Inclusion 2/4	Inclusion 4/6	Inclusion 3/8

Problem with originating inclusion??
Or did subject C take over??

Performance:

Composite Simulator Score 7/9
Penetrations 7/7

Comment: Average hours.

Members of group:

	Inclusion		Control	
	Expressed	Wanted	Expressed	Wanted
Subject A	9	6	3	2

I= Compulsively driven toward people. Difficulty being alone for extended periods of time, needs to be accepted or belong, fears rejection. Seeks out people and socializes with ease. C= "Rebel" type.

Subject B	8	5	2	1

I=Presents public image of gregarious beyond actual need for socialization and inclusion. Moderate need for companionship. C= "Rebel" type.

Subject C	4	6	4	0

I=Moderate level of social interaction and manifest flexibility. Have little concern of rejection. C= Comfortable and confident in making decisions and assuming responsibility. Need for recognition.

ANNEX D.

Example of team with low FIRO-B compatibility that performed well on AWACS (Team #12):

Reciprocal 8/8	Origin 5/6	Interchange 6/7
Control 3/5	Control 4/5	Control 3/6
Inclusion 4/4	Inclusion 4/6	Inclusion 8/8

Performance: Composite Simulator Score 4/9
Penetrations 2/7

Comment: High hours and good evaluations.

Members of group:

	Inclusion		Control	
	Expressed	Wanted	Expressed	Wanted
Subject A	0	0	2	1

I= Highly selective, avoids rejection - most comfortable moving away from people. C= "Rebel" type.

Subject B	9	9	1	3

I= Extremely drawn toward people and interacting - needs to be accepted and to belong. Doesn't tolerate living alone. C= "Rebel" type.

Subject C	9	5	8	0

I= Outwardly gregarious but more selective than B. Moderate in need for companionship. C= Makes decisions and takes on responsibilities. Expresses self-confidence and adequacy. Intense needs for recognition and maintenance of superiority.

Multi-Crew Workload Issues Onboard The Nimrod MR2 And Nimrod MRA4

Alan Felstead
Human Factors Engineer
British Aerospace Military Aircraft and Aerostructures
Warton Aerodrome, W429A Bldg
Warton, near Preston
Lancashire PR4 1AX, UK

1. SUMMARY

The Nimrod aircraft is an excellent example of how multi-crew collaboration in a complex operational system ensures the successful prosecution of a maritime patrol mission. This paper provides an overview of the roles of the Nimrod aircrew and examines how the complexity of aircrew interaction makes it difficult to measure crew workload and demonstrate that it is within an acceptable limit. The importance of ensuring that the aircrew workload is within an acceptable limit is also discussed.

2. INTRODUCTION

The Nimrod maritime patrol aircraft has been flying since 1969 in its first incarnation, the Maritime Reconnaissance Mark 1 (MR1). An updated version, the MR2, has been in service since 1981 and in 1996 the Maritime Reconnaissance Attack Mark 4 (MRA4) aircraft was chosen as its replacement. The MRA4 is due to enter service at the beginning of the next century.

The Nimrod MRA4 is a significant update to the MR2. It has completely new onboard systems, new wings and new engines. At the heart of the MRA4 flight deck is an Electronic Flight Information System (EFIS) which is based on the Airbus A340, central to which are seven LCD displays. These displays contain all the primary flight information, navigation information and onboard systems information which the pilots require. The utility systems (fuel, hydraulics, electrical, etc.) are all automatically controlled by a Utility Systems Management System (USMS) and the pilots will only need to intervene following a serious system fault. In the event of system failure the pilots are assisted by a centralised alert system which informs the pilots of the nature of the failure and automatically displays a set of remedial procedures. The operational performance of the pilots will be further enhanced by the inclusion of an Automatic Flight Control System (AFCS), a Flight Management System (FMS), a Ground Proximity Warning System (GPWS) and a Tactical Collision Avoidance System (TCAS).

The mission crew also have additional capability including a Tactical Control System (TCS) which has been developed by Boeing, an Electro-Optical Surveillance and Detection System (EOSDS) to aid target detection and a Defensive Aids Subsystem (DASS) to detect threats.

This update in technology has inevitably led to a reduction in the number of aircrew from an operational crew of 13 on the MR2 to an operational crew of 10 on the MRA4. This reduction of crew could have the potential of increasing crew workload unless it is complemented by the advanced onboard systems and by an effective Human / Machine interface.

British Aerospace Military Aircraft and Aerostructures is the prime contractor for the MRA4 and as such it is required to demonstrate that the aircraft is capable of achieving its performance objectives and is airworthy. As part of this requirement BAe MA&A must demonstrate that 'the workload imposed on the aircrew shall not exceed an acceptable level'.

It is the responsibility of the BAe MA&A Nimrod MRA4 Human Engineering team to provide a method for verifying that this workload requirement has been met. In order to do this a set of performance / workload measurement criteria must be produced, together with a means of relating these to an acceptable level of aircrew workload, at both an individual and a collaborative level.

This paper will provide an insight into the nature of crew collaboration onboard the Nimrod by examining the different roles of the crew and how they interact. This will serve to highlight one of the major difficulties in the assessment of aircrew workload onboard the Nimrod, namely the complexity of crew interaction.

3. THE NIMROD AIRCRAFT - AN OPERATIONAL OVERVIEW

The principal roles of the MR2 and the MRA4 aircraft are essentially the same and can be summarised as Anti-Submarine Warfare (ASW), Anti-Surface Unit Warfare (ASUW) and Search and Rescue (SAR).

3.1 ASW

Anti-Submarine Warfare relates to the prosecution of submarine contacts. The submarine's operating environment and modes mean that ASW detections can

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

be made either above or below the surface of the sea. This, in turn, drives the requirement for an ASW platform to have a sophisticated range of individual sensor systems, and a reliable, advanced means of co-ordinating those systems' outputs.

The principle activities involved with ASW are as follows:

- Acoustic Detection. The chief method of detecting and tracking a submerged vessel is through the use of sonobuoys which are launched into the sea. These sonobuoys are launched in patterns which are predetermined in order to provide the optimum coverage and thus the best chance of locating the submarine.

- Area Coverage. Another ASW search task requirement is the thorough and continuous radar/visual/EW cover of the assigned search area to ensure that, if a target does expose any part of his structure above the surface, the ASW platform's chances of detection are maximised.

- Recognition of Target. A target is going to appear as a very small, fleeting (radar or visual), or quiet, ambiguous (acoustic or EW) contact against a background of small, fleeting or quiet, ambiguous noise. Quickly recognising, and reacting to the presence of a target is vital to ultimate success.

- Effective Choice of Tactics. Employing the most effective tactics in response to the opportunity presented is equally vital.

- Speedy and Accurate Assimilation of Target Data. Data from the target may arrive by several means at different times. Each source will have different strengths and weaknesses and the crew captain must quickly assimilate the disparate data into a single, cohesive 'target' on which the crew can focus. The information received must be clear, concise and coherent if a correct decision is to be made.

3.2 ASUW

Anti-Surface Unit Warfare relates to the prosecution of surface shipping, both military and commercial. The military surface ship's operating environment means that it is relatively vulnerable to aircraft, and it therefore compensates by maximising the threat which it can pose to aircraft. Such units can carry a wide range of technologically advanced, long range anti-aircraft weapons system. Non-combatant vessels must rely on the anonymity and remoteness bestowed by millions of square miles of ocean as their only defence against intervention by aircraft. The combination of the need to search large areas of ocean, and the requirement to stand-off from potentially hostile contacts, drives the requirement for an ASUW platform to have one or more sophisticated, long-range search and classification tools, such as ESM and radar. Optimising the performance of, and output from, these sensors requires detailed and lengthy co-ordination.

The principle activities involved with ASUW are as follows:

- Area Coverage. One of the primary ASUW search task requirements is the security of the radar/visual/EW cover of the assigned search area to ensure that, if a target does pass through a surveyed area, the ASUW platform's chances of detection are maximised.

- Target Classification. Most radar contacts will appear similar at first sight. Quickly analysing, recognising, and reacting to the presence of a target is vital to ultimate success. Similar processing is required for EW-generated contacts, although their analysis is more objective than for radar returns.

- Effective Choice of Tactics and Speedy and Accurate Assimilation of Target Data. Clearly the data from surface unit contacts arrives from a number of diverse sources and so these activities have much in common with an ASW mission. A quick and accurate assimilation of the data, together with a suitable choice of tactics are essential for an effective mission performance.

3.3 SAR

Search and Rescue (SAR) is an activity which takes place during peace and war. It traditionally takes the highest priority in any list of tasks, and it varies in nature from incident to incident. SAR operations must be guided by 3 principles.

- Speed. A SAR platform must be able to react quickly to taskings, and to changing situations.

- Flexibility. A SAR platform must be able to meet varying demands with appropriately varying responses.

- Safety. A SAR platform must be able to monitor its own, and other's, safety, and be able to act to ensure this safety is preserved with as little detrimental impact on the task in hand as is possible.

4. THE CREW ROLES

Efficient aircrew communication and co-ordination is fundamental to the operational performance of the Nimrod.

The next three subsections will provide an overview of the roles of the MR2 aircrew, the roles of the MRA4 aircrew and the differences between the two. The knowledge of the Nimrod crew roles is crucial to understanding how the crew interact throughout a mission and therefore the variation in workload at both an individual and multi-crew level.

In order to aid visualisation of the environment in which the aircrew work, Diagram 1 shows the seated positions of the aircrew in a Nimrod MR2. The main change to seated positions in the MRA4 is in the 'mission area' of the aircraft, located between the flight deck and the galley. The revised seating positions of the MRA4 aircrew in the mission area are shown in Diagram 2.

4.1 The Nimrod MR2 Crew Roles

4.1.1 Captain.

The crew Captain has overall responsibility for the safety and tactical effectiveness of the crew and aircraft. The Captain either makes or authorises all tactical decisions during the mission. The Captain on a Nimrod MR2 may be a pilot, a navigator or an Air Electronics Officer.

4.1.2 Pilots.

Pilots on a Nimrod crew are trained to operate from both seats. The senior, or 1st Pilot is responsible to the captain for the safe operation of the aircraft. However, like all the crew members, the pilots receive full training in Maritime Operations and Target Recognition, and they have a significant tactical input. Also, as the best look-out position on the aircraft, the flight deck offers the best vantage point for visual recognition of contacts.

4.1.3 Flight Engineer.

The Flight Engineer is the third member of the flight deck crew and sits directly behind the two pilots. It is the Flight Engineer's responsibility to monitor the aircraft systems and ensure they are running correctly. This role requires an in-depth knowledge of the aircraft and its operational capability.

4.1.4 Tactical Team.

The Nimrod MR2 employs a 3-man tactical team comprising 2 navigators and an Air Electronics Officer. There are 2 navigator stations, Routine and Tactical.

The Routine Navigator is responsible for the safe navigation of the aircraft throughout all phases of flight. The Tactical Navigator, is responsible for the effective tactical employment of the aircraft and for meeting the mission objectives. The Senior, or 1st Navigator may be either the Routine or the Tactical Navigator. The Air Electronics Officer is responsible to the Captain for the effective use of the aircraft sensors and sensor operators, and for providing tactical assistance to the Tactical Navigator.

4.1.5 The 'Dry' Team.

The term 'Dry' Team refers to the 4 non-acoustic sensor operators; Radar, ESM (Electronic Support Measures), Radio and Spare. The 'Dry' Team operators will take turns in each role and the Radar, ESM and Spare will rotate roles every hour or so while the Radio operator will usually stay the same throughout a mission. One of the team will act as the 'Lead Dry' and will be responsible for the effective operation of the 'Dry' Team and for accurately reporting contact details to the Tactical Navigator.

4.1.6 The 'Wet' Team.

The term 'Wet' Team refers to the 3 acoustic sensor operators. One of the acoustic team will be the 'Lead' and will be responsible for the effective operation of the acoustic sensors and accurately reporting contact details to the Tactical Navigator. The acoustic operations are usually only employed during an ASW mission. During surface search missions the 'Wet' Team act as spare crew members to assist others and serve as lookouts on the beam.

4.1.7 Beam Lookouts.

The 2 lookout windows are located between the flight deck and the mission area. Those crew members who are not actively involved in their own task may be used as lookouts for visual recognition of contacts and as camera operators.

4.1.8 Sonobuoy Loading.

Sonobouys are used to locate underwater targets and these are stored toward the rear of the aircraft. As the sonobuoys are launched during a mission, the launchers will require reloading. The loading of sonobuoys will usually be carried out by the 'Wet' Team prior to take-off and by the Spare 'Dry' or whoever else is available during the mission.

4.2 Nimrod MRA4 Crew Roles.

The crew role responsibilities for the MRA4 are still in the process of being refined, however the roles of the 10-man Nimrod MRA4 crew are essentially as follows:

4.2.1 Captain.

The Captains role is the same as it was on the MR2. The Captain on a Nimrod MRA4 may be a pilot or a TACCO.

4.2.2 Pilots.

The pilots on the MRA4 will have the same duties as on the MR2 plus the additional task of routine navigation during the outbound and inbound phases of the mission. Navigation throughout the tactical phase of the mission will be performed by the TACCOs (see below). The pilots may also be required to carry out some systems reconfiguration in the event of equipment failure although routine systems management is fully automated.

4.2.3 Tactical Team.

The Nimrod MRA4 will employ 2 Tactical Controllers (TACCOs). The senior TACCO, also known as TACCO 1, will be responsible to the Captain for the effective tactical employment of the aircraft and for meeting the mission objectives. The assistant or TACCO 2 will be responsible to TACCO 1 for the effective use of the aircraft sensors and sensor operators, and for providing tactical assistance as required.

4.2.4 The Communicator.

The communicator will be responsible for the setting-up and allocation of all radio communications systems. The Communicator will also be responsible for the appropriate and timely dissemination of messages coming into the aircraft, and transmission of messages leaving the aircraft. In addition the Communicator will manage the data link and satellite communication systems under the TACCO's overall direction.

4.2.5 The 'Dry' Team.

The 'Dry' Team on the MRA4 will comprise of 1 Radar, 1 ESM and 1 spare. The operators will rotate roles as they currently do on the Nimrod MR2. The 'Dry' team will be responsible for the effective operation of their sensors and for accurately reporting contact details to the TACCOs. The ESM operator's position will also serve as the starboard beam lookout.

4.2.6 The 'Wet' Team.

The 'Wet' Team on the MRA4 will comprise of 2 acoustic sensor operators. Their duties will be much the same as they were on the MR2. It is also intended that the TACCO 2 operator will assist the 'Wet' Team and therefore act as a supplementary member.

4.2.7 Beam Lookout.

As described above, the ESM operator's seat will double as the starboard lookout. The port lookout will be manned when required by any operator not actively involved in another task.

4.2.8 Sonobuoy Loading.

Although the loading activity will be similar to the MR2, the MRA4 will carry more sonobuoys and this will be co-ordinated by the TACCO 1.

5. DIFFERENCES BETWEEN AIRCREW COMPOSITION ON THE MR2 AND THE MRA4

Clearly the crew roles have changed as a result of the aircraft update. The principal issues concerned with this change are:

5.1 In the Flight Deck

- There is no longer a Flight Engineer. The monitoring of the utility systems is now carried out automatically by a Utility Systems Management System (USMS).
- The pilots have taken on the additional role of Routine Navigation role during the outbound and inbound phases of the mission.
- The performance of the pilots has been enhanced by an Electronic Flight Information System (EFIS), a Flight Warning System (FWS), an Automatic Flight Control System (AFCS), a Flight Management System (FMS), a Ground Proximity Warning System (GPWS) and a Tactical Collision Avoidance System (TCAS).

5.2 In the Mission Area

- The 'Wet' Team has been reduced by 1 but can expect the TACCO 2 to help out during ASW missions.
- The Tactical team has been reduced by 1.
- The Radio or 'Comms' operator is no longer a member of the 'Dry' Team.

- The reduction in mission crew has been compensated by the additional capability resulting from such systems as a Tactical Control System (TCS), an Electro-Optical Surveillance and Detection System (EOSDS) and a Defensive Aids Subsystem (DASS).
- An enhanced cabin layout has resulted in improved unaided communication between crew members.
- Throughout the aircraft Human Factors specialists have worked closely with the system designers and operators. This has resulted in an optimum HMI that should improve crew performance and ensure the reduction in aircrew has not had an adverse effect.

6. THE REQUIREMENT TO MEASURE CREW WORKLOAD

It is important at this point to introduce the requirement to measure the Nimrod crew workload and demonstrate that it is within an acceptable level.

The new crew composition was arrived at as the result of several predictive workload analysis and task analysis activities. These analyses concluded that, despite the reduction in overall aircrew members, the new aircrew composition, together with the enhanced onboard systems, would increase the effective performance of the Nimrod. However this predictive analysis does not remove the requirement to actually demonstrate that the crew workload on the MRA4 is within 'an acceptable level'.

Maintaining an acceptable crew workload is important both from a mission effectiveness and from a safety viewpoint. The Nimrod, whether it is an MR2 or an MRA4, is constructed and equipped to collect large amounts of data from various sources using diverse means. In order to function effectively, **the Nimrod crew must communicate both facts, and ideas, quickly and effectively**. This is the 'key' to effective crew performance. From a mission effectiveness viewpoint it is performance of the Tactical Navigator (or TACCO 1 in the MRA4) that is central to the success of the mission. It is the Tactical Navigator / TACCO 1 who must make the tactical decisions and who must be able to enforce those decisions. If the workload of the crew, and the Tactical Navigator / TACCO 1 in particular, becomes unacceptably high then the effectiveness of the entire mission will be impaired

The Nimrod has long mission times, often well over 8 hours. During the tactical phase of the mission much of the time will be spent at heights of only 200 feet above sea level. Crew situational awareness, especially in the flight deck, is therefore extremely important from a flight safety viewpoint, and the safety implications of exceeding an acceptable workload are obvious.

Thus we have seen the need to measure crew workload and why it is essential to ensure that it will remain within an acceptable level. The next section will examine one of the biggest difficulties in measuring Nimrod aircrew workload; the complexity of crew interaction.

7. COMPLEXITY OF CREW INTERACTION

This section uses the Nimrod MRA4 to illustrate aircrew interaction during a typical mission. Due to the difference in crew composition, Nimrod MR2 crew interaction will obviously differ from this but the level of interaction is much the same. Similarly the interactions between the aircrew will depend on the type of this mission; the example presented here is more akin to an ASW mission.

Diagram 3 is a simplistic overview of crew interaction during a typical ASW mission onboard the Nimrod MRA4. The diagram shows the various teams in which the crew work and the interactions which are at both an intra-team and inter-team level. The crew role interactions shown on this diagram can be summarised as follows.

7.1 On the Flight Deck.

The Non-flying Pilot will assist the Flying Pilot by carrying out navigation, flight management and general maintenance duties. Both of the pilots will be involved with the handling of remedial procedures following receipt of an alert. In addition to the interaction between the pilots, information will be passed to and from the mission crew. This information will usually be of a tactical nature which will impact on the mission.

7.2 The Tactical Team.

The central co-ordination of the mission lies within the tactical team and this is clearly illustrated in diagram 3. As can be seen from diagram 2, TACCO 1 is located in the centre of the mission area and will co-ordinate the mission by receiving information and passing orders to the rest of the crew. The TACCO 2 will assist TACCO 1 in setting up the tactical control and display systems, will ensure the appropriate sonobuoy release selections are available and will offer input/opinion on the choice of tactic. The TACCO 1 operators ability to make decisions and ensure they are acted upon is the key to a successful mission.

7.3 'Dry' Team.

The Radar, ESM and Spare operators will work together to provide information to, and receive orders from, the Tactical Team but will also have contact with

other crew members. The ESM, for example, has the additional task of starboard beam lookout and as such will also be communicating with the pilots. The Spare may act as a sonobuoy loader during ASW missions or may act as the port beam lookout.

7.4 'Wet' Team.

The 2 acoustic operators will work together and Acoustic 1 will act as the lead passing information to, and receiving orders from, the Tactical Team. TACCO 2 can be expected to help out as a supplementary 'Wet' Team member. The pilots will pass on received sonobuoy signal information to the 'Wet' team. During an ASUW mission especially, either acoustic operator may be acting as a port beam lookout.

7.5 Comms.

The Comms operator will pass radio messages, both to the tactical team and to the flight deck and will send out radio messages which have been composed by the pilots and by the TACCOs. The Comms operator will not usually directly assist the 'Dry' team but is located between the 'Dry Team and TACCO 1 so this cannot be ruled out.

7.6 Outside World.

Not shown in Diagram 3, but in addition to the above internal communications, the Comms operator, the pilots, the TACCOs and the Radar operator will all have some communication with the outside world.

8. INFORMATION FLOW / DECISION MAKING

Diagram 3 does not show every possible interaction between the crew; it describes typical interactions, based on the individual duties of each of the aircrew. In reality the roles of the Nimrod aircrew will be far more fluid and interaction is likely between all parties at some point in the mission. Nevertheless, the diagram does help to illustrate the complexity of the crew structure.

To further illustrate the complexity and flexibility of the aircrew roles, Table 1 shows examples of the type of data passing between the operators.

The advancement in technology in onboard systems has increased the ability to flow information between the operators. On the MR2 only the Tactical Navigator received the full tactical picture and so had to be relied on to make the correct decision. On the MRA4 the operators have the ability to view each others displays and so are able to form their own opinions. Even the pilots have the ability to make tactical assessments because they will receive a tactical display in the flight deck. While it is useful for all operators to be able to send advice on tactics to the TACCO 1 operator this increased capability could undermine the central decision making role of the TACCO 1. It must be ensured that this does not happen on the MRA4 and that the TACCO 1 operators decision on all tactical matters is final.

It can also be seen from Table 1 that, although the crew members have pre-defined roles, the activities that each member will be carrying out will vary considerably, depending the scenario, and cannot be fully defined in advance. Again the new MRA4 systems increase the ability for operators to take over each others activities and whilst this will improve the performance of the crew as a whole it makes individual crew performance and workload even harder to measure.

9. CONCLUSION

It is the complexity and flexibility of Nimrod aircrew interaction which makes both individual and collaborative crew workload measurement difficult and an acceptance level very hard to place. The tasks an individual operator will be performing may vary, depending on the scenario and are hard to predict in advance. On the MRA4 it is even easier to take over each others tasks. How is it possible to measure how well an individual performs a set of tasks when the precise tasks that the individual will be performing cannot be defined in advance ?

A further problem which adds to the difficulty of assessing workload levels is that the crew do not even stay in a fixed position during the mission. During a typical mission the aircrew will be moving about the aircraft; they will be cooking food in the galley, eating and taking rest periods at irregular intervals. This makes the taking of measurements at pre-defined positions, such as the mission crew workstations, very difficult.

An 'acceptable workload level' is always very hard to define. During certain missions some of the crew will have periods when they have no duties at all to perform. Does this mean that they are being under utilised ? It is British Aerospace's responsibility to demonstrate that the crew workload never becomes unacceptably high but the crew members routinely assist each other so exactly when is the workload on an individual too great ?

The problems illustrated in the last three paragraphs are all issues associated with the measurement of multi-crew workload and this is not the first time they have been encountered. Nevertheless, the Nimrod aircraft provides an excellent practical example of a multi-crew environment where it is essential to ensure an acceptable level of workload is not exceeded.

FURTHER READING

For further information regarding the actual techniques chosen for the measurement of aircrew workload onboard the Nimrod, the reader is guided to the paper by Steve Harmer entitled 'Multi-Crew Workload Measurement for Nimrod MRA4' which may also be found in these proceedings.

ACKNOWLEDGEMENTS

The author would like to thank Squadron Leader Andrew Steel for his invaluable help in providing operational information for this paper.

Diagram 1. The Nimrod MR2 Crew Position Layout

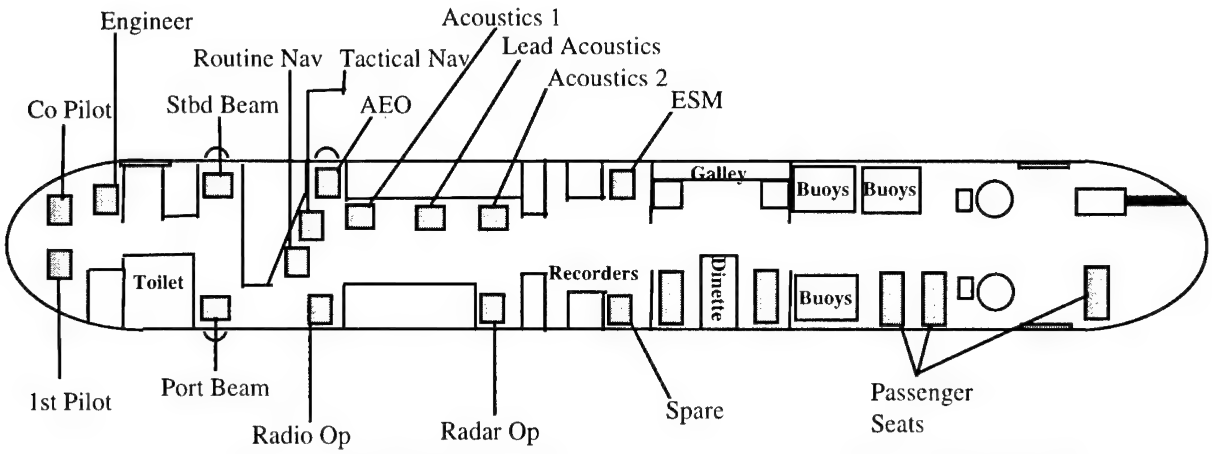


Diagram 2. The Nimrod MRA4 Mission Area Crew Position Layout

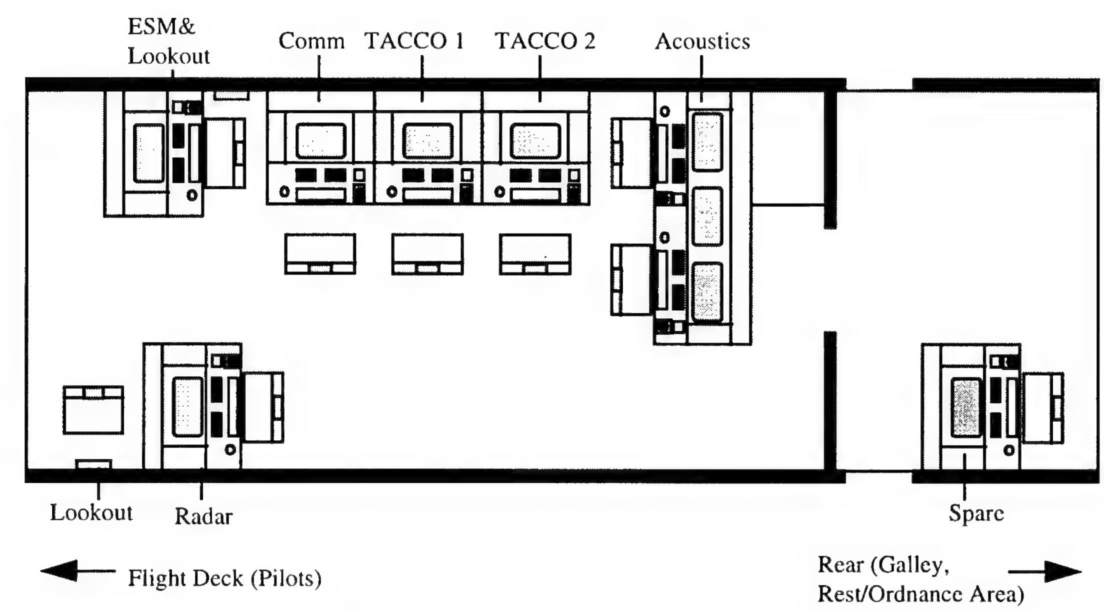


Diagram 3. Example of Typical Nimrod MRA4 Crew Interaction during an ASW Mission

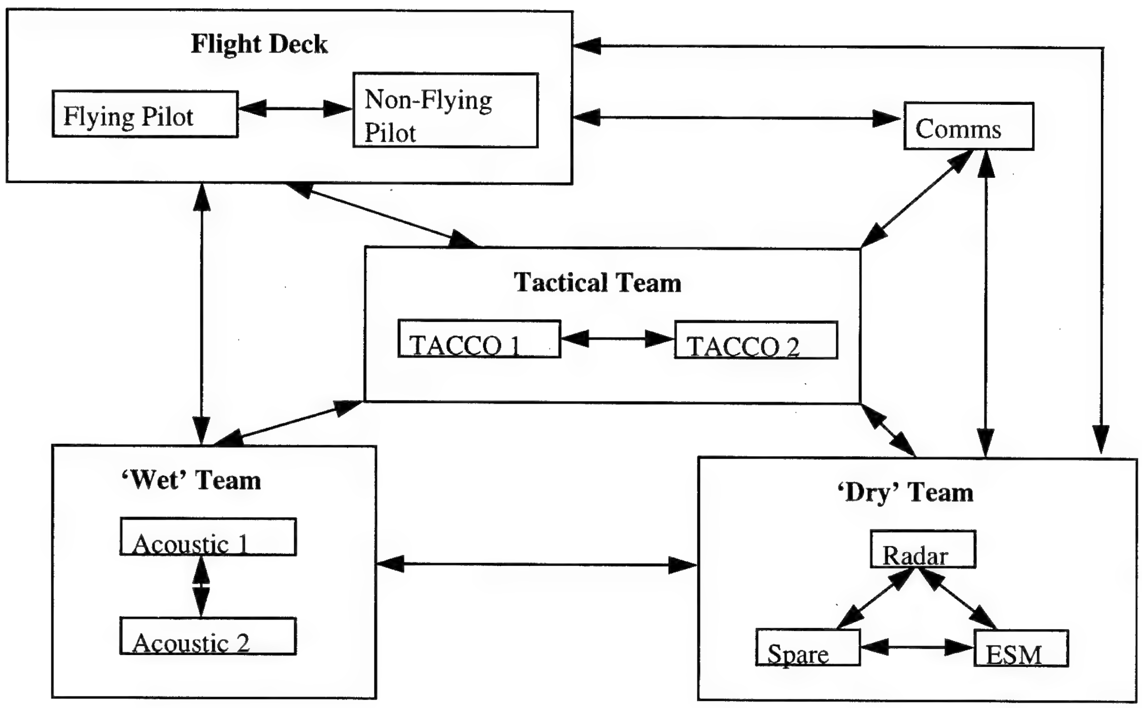


Table 1. Data Flow Examples

Type of Data	Sent By	Received By	Examples
Tactical Decisions	TACCO 1	All Operators	Decision to prosecute a specific target.
Tactical Directions	TACCO 1, TACCO 2	All Operators	Instructions to prosecute a specific target using specific tactics.
Tactical Ideas	All Operators	All Operators	Suggestions to aid tactical decision making (e.g., "Target appears to be speeding up")
Flight Safety Decisions	Pilots	All Operators	Decision to abort sortie due to a lightning strike.
Flight Safety Data	All Operators	Pilots	Visible evidence of a lightning strike on the airframe.
Contact Data	'Wet', 'Dry', Comms	TACCO 1, TACCO 2	Parametric, classification, position and/ or behaviour / activity reports.
System Requirements	TACCO 1, TACCO 2	'Wet', 'Dry', Comms	Instructions on which modes of sector operation to use.
System Status Data	'Wet', 'Dry', Comms	TACCO 1, TACCO 2	Reports on modes in use, success of those modes and overall system performance.
Beam Lookout Information	Beam Lookouts	All Operators	Tactical and flight safety reports (e.g., identity of ships, sea state, etc.)

Multi-Crew Workload Measurement for Nimrod MRA4

Steven Harmer
Human Factors Scientist
British Aerospace Sowerby Research Centre
FPC 267, P.O. Box 5
Filton, Bristol BS12 7QW, UK

Summary

British Aerospace, as Prime Contractor for the UK's Replacement Maritime Patrol Aircraft, the Nimrod MRA4, due to enter service at the beginning of the next century, is required to demonstrate that the crew workload levels associated with operating the aircraft do not exceed acceptable levels. In order to do this BAe must be able to define this acceptance level and provide a meaningful mechanism for measuring workload in a multi-crew environment, where task allocation is highly dynamic and team working is essential.

To achieve this BAe has adapted two existing 'individualistic' subjective workload assessment techniques, (the NASA Task Load Index and Instantaneous Self Assessment), for the measurement of multi-crew workload. Together with objective crew performance measures, the techniques have been used to measure current Nimrod crew workload levels to establish a baseline against which Nimrod MRA4 levels will be compared.

This paper describes the techniques, the method proposed for comparing between two different crew compositions and the issues associated with deriving crew workload acceptance criteria.

1. Introduction

British Aerospace is currently developing the Replacement Maritime Patrol Aircraft (RMPA) for the Royal Air Force, under contract to the UK Minstry of Defence. This will be a much improved version of the Nimrod Maritime Reconnaissance (MR) aircraft, which has been in service since 1969. The current version, the MR2, has been operated by the RAF since 1981 and its replacement, designated the Maritime Reconnaissance Attack Mark 4 (MRA4) is due to enter service at the beginning of the next century.

As Alan Felstead has described in his paper contained in these proceedings, the MRA4, although sharing the same basic air vehicle as its predecessor, will incorporate many new onboard systems, new wings and new engines. In the flight deck, the significantly upgraded cockpit technology will include an Electronic Flight Information System (EFIS) based on the Airbus A340 system, a new Utility Systems Management System (USMS), an advanced Flight Management System (FMS), a Ground Proximity Warning System (GPWS) and a Tactical Collision Avoidance System (TCAS).

The mission system will incorporate a Tactical Control System (TCS), developed by Boeing, integrating a sensor suite which includes new and advanced acoustics, Radar, and Electronic Surveillance Measures (ESM) systems, plus the addition of an Electro-Optical Surveillance and Detection System (EOSDS), a Datalink communications system and a sophisticated Defensive Aids Subsystem (DASS).

In parallel with updates to the air vehicle and its mission system there have been changes in the crew composition and overall numbers. The MR2 operational crew of 13 has been reduced to 10 for the MRA4, with the removal of the Flight Engineer from the cockpit, a reduction from 3 to 2 in the acoustics (or 'wet') team, and the removal of the Air Electronics Officer (AEO).

The new crew will be 2 pilots, 2 Tactical Coordinators (TACCOs), a Communicator, 2 acoustics sensor operators and a 'dry' team comprising a Radar operator, ESM operator and a 'spare' whose duties include the loading and setting of sonobuoys. As in previous Nimrod crews the dry team will rotate to provide an opportunity for Radar and ESM operators to rest during long operations.

As Alan Felstead described in his paper, when compared to the existing Nimrod MR2, there will be a number of task reallocations across the crew and between the crew and the new automated systems. To ensure that these reallocations match the capabilities and limitations of the new crew configuration, crew workload has been, and will continue to be, assessed throughout the aircraft development.

2. RMPA/Nimrod MRA4 Workload Assessment to Date

Workload assessment for the RMPA started before British Aerospace was selected as Prime Contractor for developing the new Nimrod MRA4. Prior to the competition for the contract, the UK Defence Evaluation Research Agency (DERA) commissioned a study by Canadian Marconi Corporation (CMC) to look at the crew requirements in terms of both numbers and composition necessary to meet the overall system performance requirement. As with many replacement military systems today, a reduction in manpower was sought together with an increase in the effectiveness of the system. This put crew workload high on the risk agenda.

The approach taken by CMC was to model the current Nimrod MR2 crew workload for a composite Anti-Submarine Warfare (ASW)/Anti-Surface Warfare

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

(ASuW) mission using a proprietary workload prediction tool known as System Operator Loading Evaluation (SOLE) which is based upon the McCracken and Aldrich VACP (Visual, Auditory, Cognitive, Psychomotor) workload modelling technique. Having modelled the MR2 crew workload, using data collected during simulation runs at the Nimrod training facility at RAF Kinloss, CMC then produced a model of RMPA crew workload based upon the assumed crew roles and the required sensor fit. Comparisons were then made between the MR2 and MRA4 crew workload profiles in order to identify workload 'hotspots' which were used to refine the RMPA requirement.

Prior to the release of the results of the CMC analysis, British Aerospace and Boeing had also performed a similar, though less comprehensive study of crew workload, for their proposed Nimrod 2000 (later to be designated the MRA4). Like the CMC study, BAe and Boeing used current Nimrod crew workload levels as a baseline from which to project future MRA4 crew workload levels. The aim of the crew workload prediction at this time was to assess options for the allocation of tasks amongst the new smaller crew complement and to identify potential workload risks associated with the proposed system configuration.

Having been awarded the Nimrod MRA4 contract, BAe and Boeing have continued to monitor crew workload levels during the development of the air vehicle and mission systems. Now attention is turning towards the *demonstration* of acceptable crew workload levels during aircraft acceptance trials, and it is on the techniques chosen for this activity which the remainder of this paper will concentrate.

3. Techniques Chosen for Workload Assessment

A combination of techniques is being used for the assessment and demonstration of crew workload levels during the acceptance trials for Nimrod MRA4. These fall into two categories: *objective* performance measures, and *subjective* workload assessment measures.

The approach taken to demonstrating that crew workload is acceptable is to assess firstly, that crew performance meets an acceptable level and secondly, that in meeting that performance level, individual workload does not exceed an acceptable level.

3.1 Objective Crew Performance Measurement

Crew performance is being assessed using primary task crew performance measures, or 'Crew Measures of Effectiveness' (MoEs), which have been derived from a detailed analysis of two multi activity test (MAT) missions, one ASW and one AsuW, and one dedicated flight deck sortie into which a number of system failures (which result in emergency conditions) are injected. Working closely with company Subject Matter Experts and DERA personnel at Boscombe Down, an agreed set of crew MoEs have been established which are largely independent of the system being used and therefore can be used for the comparison of crew performance for the current Nimrod MR2 and the new Nimrod MRA4. These are a combination of anecdotal qualitative 'measures' - did the crew perform an activity in accordance with procedures, was the tactic employed appropriate for the situation etc. - and quantitative measures (i.e., times and accuracies). Table 1 shows an example of some of the MoEs being used. It is not the intention of this paper to discuss the details of how these crew performance measures were derived, needless to say, Subject Matter Expert involvement and a thorough understanding of the test missions were fundamental to the process. Instead, the remainder of the paper will concentrate on the subjective workload assessment techniques chosen, and the tools used to support these techniques.

3.2. Subjective Crew Workload Measurement

Two complementary subjective workload assessment techniques are being used to assess crew workload levels experienced during the two MAT missions and the flight deck 'emergencies' sortie. These two techniques are the NASA Task Load Index (TLX) (ref. 1) and a modified version of the Instantaneous Self Assessment (ISA) technique. The latter was originally developed by the National Air Traffic Services (NATS) for the assessment of air traffic controller workload in the development of new Air Traffic Control systems. The technique can be used for the on-line assessment of operator workload and recent studies (ref. 2) have shown that the technique is less intrusive than other popular on-line assessment techniques such as SWAT (Subjective Workload Assessment Technique) (ref. 3.).

The NASA Task Load Index

The NASA-TLX has been used widely for the evaluation of operator workload in many military and civil applications. It requires subjects to rate tasks against the following six workload constructs:

- Mental Demand
- Physical Demand
- Temporal Demand
- Effort
- Own Performance
- Frustration

The technique is being used to assess subjective workload in each phase of the two MAT missions and during each of the failure conditions in the flight deck emergencies sortie. During simulated runs of the test missions, crew members are asked to provide ratings at the end of each phase of a mission. The ASW mission (MAT1) has, for example, 14 phases starting with Take-Off and Climb, and continuing through the 'Join Task Group' phase, Sonobuoy Barrier Deployment, Target Detection, Target Localisation etc., and lasts for approximately 6 hours. The points at which TLX ratings are provided are predetermined and carefully planned to avoid interference with crew primary task performance. Ratings are provided against each of the 6 workload constructs on rating sheets contained in a booklet provided to each crew member before the start of each

test run. Following the test mission, all crew members are required to perform pairwise comparisons of the 6 workload constructs, to enable weighted workload scores to be calculated for each crew position in each mission phase.

Whilst the NASA-TLX offers an easy-to-administer subjective crew workload assessment technique, capable of providing a degree of diagnosticity for the sources of the workload experienced, e.g. time pressure, mental demands, frustration etc., its implementation means that workload ratings are aggregated over phases which can be as much as 30 minutes long. The results obtained therefore provide a coarse profile of a crew member's workload throughout the mission. To overcome this, and to provide a finer granularity of workload ratings, the ISA technique is used in parallel.

Multi-Crew Instantaneous Self Assessment

Multi-Crew Instantaneous Self Assessment (MC-ISA) is British Aerospace's derivative of the ISA technique, originally developed by the National Air Traffic Services to investigate Air Traffic Controller workload. The technique involves subjects providing a rating on a simple five-point scale when cued. The simple scale ranges from Very Low through Fair to Very High, with intermediate levels of Low and High. The ratings are provided via a simple interface button box with 5 coloured buttons which represent the five workload levels. The interface box also incorporates two small Light Emitting Diodes (LEDs) which flash to cue the subjects to provide a rating. For the Nimrod MRA4 MATs, the LEDs flash for 60 seconds every 4 minutes and subjects are briefed to provide a rating during the 60 second cueing period. Subjects are also told to provide a rating only when their primary task will allow, and not to interrupt their tasks solely for the purpose of making a rating, the aim being to minimise the intrusive effects of the rating task. In practice we have found that subjects rarely miss ratings, are able to respond quickly to cues and consider the intrusive effects to be minimal. Response times can be measured, although use of these times as secondary performance measures is not proposed. Subject cueing and the collation of ratings and response times is centrally controlled by a personal computer and therefore the assessment technique is simple to administer and the results easy to analyse.

Unlike the NASA-TLX, MC-ISA uses a simple unidimensional rating scale, which provides a single rating for 'overall' subjective workload. Whilst lacking the diagnostic qualities of the TLX it provides a mechanism for obtaining frequent ratings throughout a long mission without the need to interrupt a subject's 'work-flow'. When both techniques are used together, they complement each other to provide an effective method for assessing the workload levels experienced by a crew and the sources of the demands which contribute to that workload.

4. Determining Workload Acceptance Levels

As with earlier predictive workload assessments for RMPA and Nimrod MRA4, the demonstration of acceptable crew workload levels is based upon a comparison with Nimrod MR2 levels. Using the techniques described, crew workload 'baselining' trials were performed at RAF Kinloss in December 1997 to establish a baseline dataset to support comparisons during Nimrod MRA4 acceptance trials. The approach that will be taken to demonstrating that crew workload is 'acceptable', will be to assess firstly, that the effectiveness of the Nimrod MRA4 System (including the crew) meets an acceptable level, and secondly, in meeting that performance level, individual crew member subjective workload does not exceed an acceptable level (see figure 1).

For both crew/system effectiveness and subjective workload, acceptance criteria are based upon a comparison with the Nimrod MR2 baseline dataset. In addition, crew/system effectiveness will also consider the specified system performance requirements for MRA4, which in many cases significantly exceed the performance of MR2.

Comparison of subjective crew workload measures is complicated by the reduction in crew complement and changes in crew composition. A number of the crew positions can be compared one-for-one, for example: the ESM operator, Radar Operator and Communications Operator. However in the Flight Deck, Wet Team and Tactical Team, there have been reductions in the crew complement and there is no longer a one-to-one mapping with Nimrod MR2. Comparisons here will be made at a 'Functional Team' level (see table 2) by combining workload scores and calculating a mean workload level for the functional team with a standard deviation. The aim for MRA4 will be to maintain or reduce the mean functional team workload level whilst maintaining a low standard deviation within the team. Comparisons made in this way overlook any differences in the actual tasks performed by individual crew members, and simply compare the workload levels experienced by the crew regardless of the changing nature of their jobs which may result from task reallocations, increases in system automation or additional responsibilities. At first, this may appear to be a weakness in the approach, but it is argued that the detailed tasks that individual crew members perform are of less interest, so long as the overall crew performance is acceptable <u>and</u> that workload levels experienced by individuals are acceptable and balanced throughout the crew.

During Nimrod MRA4 system development, where individual workload levels appear excessively high or where an imbalance in crew workload occurs, changes to the allocation of task or changes to operating procedures may well provide a solution before changes in the system design, such as increasing automation levels, need to be considered.

5. Discussion

The approach described incorporates the use of 'individualistic' workload assessment techniques, applied in a multi-crew environment. As with all workload assessment techniques, these have both strengths and weaknesses. On the positive side, both the NASA-TLX and MC-ISA are simple to administer, requiring minimal subject training. This has allowed BAe to perform baselining trials at RAF Kinloss involving three full Nimrod MR2 crews, over a two week period, using simulation facilities which are in high demand, without significant disruption to operations or to training. The results obtained are easy to analyse and, together with video data, high workload tasks can be easily identified for further detailed analysis.

The ISA system, initially designed to look at single operator workload, can be readily adapted to investigate multi-crew workload. Although the technique uses a simple unidimensional rating scale, it is effective at identifying workload hot-spots within the crew. Its potential as a secondary task performance measurement system has yet to be explored, but an initial inspection of crew response times suggests a correlation exists between high ratings and high response times.

The NASA-TLX, although useful for identifying the sources of demand associated with crew workload, requires a detailed post mission debrief to investigate the true sources of high ratings. With a crew of 10 (Nimrod MR2), or 13 (Nimrod MRA4) performing missions often in excess of 7 hours, the time needed for detailed debriefing can present a problem.

However, the greatest problem associated with the use of the individualistic techniques chosen does not relate to their application to the assessment of multi-crew workload. Defining an acceptable crew workload level, based upon comparison with an existing system, involves the same problems for a multi-crew system as with a single operator system, e.g:

- *How do we overcome the subjectivity of results obtained ?*
- *How do we ensure that users of a new system are sufficiently familiar with the operation of that system to perform meaningful comparisons with their existing systems ?*
- *If workload is higher for the new system, how much higher is truly unacceptable ?*

These problems represent a challenge to the Nimrod MRA4 developers and to the science of workload assessment generally. If we are able to prove that workload does not exceed acceptable limits, we must be able to define these limits in quantifiable terms. However, it is the author's personal opinion that the workload 'red-line' as it is known, cannot be defined with sufficient accuracy and that the assessment of operator workload will always rely on the subjective assessment of experts.

6. Conclusion

Quantifying the workload levels for a crew performing the complex tasks associated with operating a maritime patrol aircraft is a considerable challenge. To then demonstrate that these levels are truly *acceptable,* with any degree of confidence, one could argue is impossible. However British Aerospace and Boeing are committed to developing an aircraft which can be operated safely and effectively by the intended user population. In doing this it has been necessary to develop a pragmatic approach to crew workload measurement which, if it cannot provide a definitive indication of acceptable crew workload levels, can at least provide a 'richer picture' of the effectiveness of the crew/system design and the demands placed on individual operators.

7. References

1. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), Human mental workload. Amsterdam: Elsevier.

2. David, H., & Pledger, S. (1995). Intrusiveness of n-line self-assessment in ATC simulation using keyboard and speech recognition. EEC Report No. 275, sponsored by Eurocontrol. Directorate of EATCHIP Development.

3. Reid, G. B., Eggemeier, F., & Nygren, T. (1982). An individual differences approach to SWAT scale development. In Proceedings of the Human Factors Society 26th Annual Meeting (pp. 639-642). Santa Monica, CA: Human Factors Society.

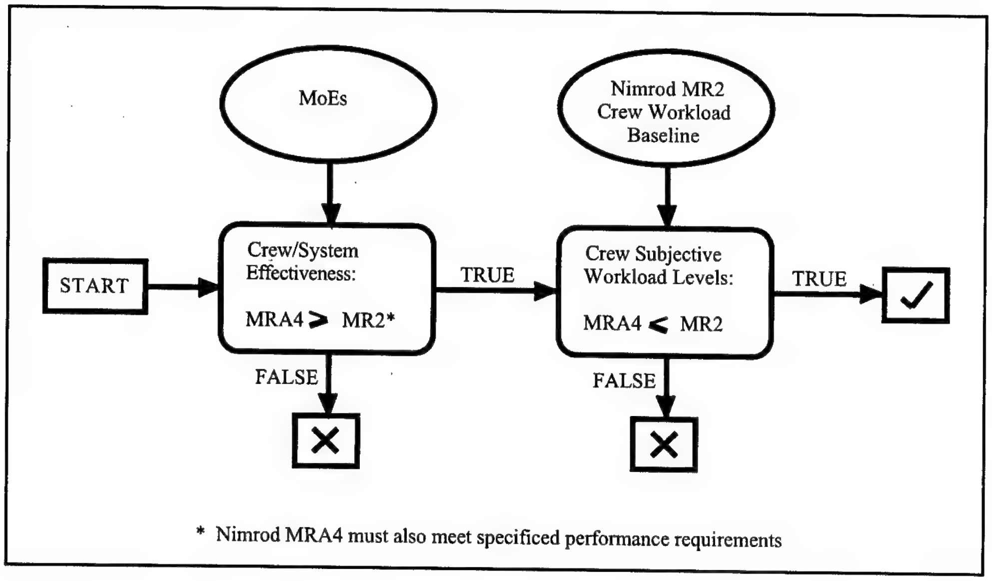


Figure 1. Overview of Nimrod MRA4 Crew Workload Acceptance Methodology

MoEs for ASW MAT1 Mission Event "Gain Contact on Target Submarine"		
MoE	**Measurement Technique**	**Information Required**
Time taken to detect presence of target	Stopwatch/Video Analysis	Time from contact available to operator noting presence
Time taken to report contact to TACCO	Stopwatch/Video Analysis	Time from contact observed to reported
Range of target from buoy when contact gained	Observation/Data Analysis	Range from target to buoy when detected
Time taken to report contact to ASWACU	Stopwatch/Video Analysis	Time contact observed to message transmitted
Were follow up reports made in accordance with Optask ASW	Observation/Data Analysis	Y/N & times between reports (to nearest minute)
Accuracy of each transmitted position	Observation/Data Analysis/Video Analysis	Data transmitted vs actual target position

Table 1. Example of Crew Measures of Effectiveness (MoEs)

Nimrod MR2 - MRA4 Crew Comparisons		
Functional Team	**Nimrod MR2**	**Nimrod MRA4**
Flight Deck	Pilot 1 Pilot 2 Flight Engineer	Pilot 1 Pilot 2
Tactical Team	Tactical Navigator Routine Navigator Air Electronics Officer	TACCO 1 TACCO 2
Wet Team	Lead Acoustics Operator Acoustics Operator 1 Acoustics Operator 2	Acoustics Operator 1 Acoustics Operator 2
Dry Team	Radar ESM Spare Dry/Ordnance Communications	Radar ESM Spare Dry/Ordnance Communications

Table 2. Use of 'Functional Teams' for Comparing Nimrod MR2 and Nimrod MRA4 Crews.

ASSESSING OPERATORS' POTENTIAL FOR COLLABORATION IN COMPLEX SYSTEMS

R.E. King
United States Air Force Research Laboratory
Collaborative Systems Technology Branch
AFRL/HECI, 2255 H. Street, Bldg. 248
Wright-Patterson AFB, OH USA 45433-7022
1-937-255-7581
rking@al.wpafb.af.mil

J.D. Callister
Department of Defense Medical Support
Brooks AFB, TX USA

P.D. Retzlaff
University of Northern Colorado
Greeley, CO USA

SUMMARY: Just as the specific *context* of a complex situation must be appreciated, we need to understand the individual characteristics of the specific human *actors* in a given complex system. The *Armstrong Laboratory Aviator Personality Survey* studies individual differences and their contribution to team functioning. The Crew Interaction scales of ALAPS include: *Deference, Dogmatism, Impulsivity, Organization, Risk Taking, and Team Orientation.* Preliminary data collection of candidates for USAF undergraduate pilot training (N = 1131, 124 of which are female) with an average age of 22.6 (sd = 2.9), resulted in a full range of item endorsement. Variance in potential for successful collaboration in pilot and astronaut candidate populations is of particular interest to researchers of team performance since these candidates have all been thoroughly screened for cognitive functioning (Carretta, Retzlaff, Callister, & King, 2) and have also proven themselves to be highly technically competent. Nevertheless, experience suggests some of these otherwise highly qualified individuals are destined to ultimately fail in a collaborative environment. The concept of the "Right Stuff" may need to be expanded to include successful *team* performance, due to the demands of global engagement and long-term space exploration.

TEXT: Today's, and tomorrow's, successful systems operator may be very different from the barnstormers and aces of yesteryear. In particular, military aviation has changed from the days of dog fighting to modern multi-crew, highly coordinated missions. Due to ever-increasing levels of complexity, future military and space operations will be more highly dependent on team functioning. In this environment, high levels of extraversion and conscientiousness in addition to traits of organization and team orientation, and, to an extent, risk taking, may be very valuable. Conversely, impulsivity, dogmatism and authoritarianism may be deleterious to the accomplishment of the mission. Military aviation is increasingly an interpersonal endeavor. A study of the psychological factors leading to success in aviation will enable researchers to select-in the most psychologically and cognitively suited person for the mission rather than only select-out those who are inappropriate. The need is rapidly changing from individual excellence to group excellence, which may be more than the mere sum of the characteristics of individual members of a team. Even individuals who have done well in solitary pursuits may fail when placed in an interpersonally demanding situation. Space exploration serves as a useful model. Due to lengthening space missions and the increasingly interpersonal nature of space exploration, combined with an expanding multicultural flavor, selection of individuals with desirable interpersonal qualities is becoming paramount. The basic question in the back of the minds of evaluators for the space program is: "Could I spent six months in a bathroom with this person?"

United States Air Force pilots, and pilots from other services and nations, are selected on instruments heavily loaded on general intellectual ability (Carretta, Retzlaff, Callister, & King, 2). Special programs often require additional selection methods, such as psychological testing and structured interviews. Moreover, the psychiatric standard of fitness (1), which is an intrapsychic phenomenon, is yielding to a psychological model of suitability, which is more context specific.

ARMSTRONG LABORATORY AVIATOR PERSONALITY SURVEY (ALAPS):

The Armstrong Laboratory Aviator Personality Survey (ALAPS, 4), is designed to assess the psychological and psychiatric characteristics and concerns of entry level female and male pilots. The ALAPS is currently administered during the Enhanced Flight Screening (EFS) program (3) at the United States Air Force Academy and Hondo Airport, Texas.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

The development of the ALAPS was accomplished as follows: 15-18 dimensions of interest were identified, including both select-in and select-out domains. An initial pool of 24 items per scale was written. A sample of about 100 college students took an initial form to identify statistically poor items. Those items were replaced and the initial form was administered to 300 pilot candidates. A final set of 16 items per scale was retained for the final form. Entire scales which failed to reach appropriate levels of reliability were eliminated.

Table 1. Scales of ALAPS

Validity
1. Reliability
2. Disclosure
3. Intra-individual consistency

Personality
1. Aggressiveness
2. Confidence/ Narcissism
3. Negativity/ Passive-aggression
4. Order/ Compulsivity
5. Socialness

Psychopathology
1. Affective Lability
2. Alcohol Abuse
3. Anxiety
4. Depression

Work Styles/ Crew Interaction
1. Deference/ Submissiveness
2. Dogmatism/ Authoritarianism
3. Impulsivity
4. Organization
5. Risk Taking
6. Team Oriented

Due to the focus on collaborative systems of this paper, only the Work Styles/Crew Interaction scales of the ALAPS will be delineated:

Deference: High scorers are deferent to a fault. They are submissive and quiet. They concentrate on their job and are not likely to question the status quo.

Dogmatism: High scores believe they are always correct and are not open to change. They are interpersonally authoritarian. They are intolerant of other people, their ideas and their actions.

Impulsivity: High scorers act first and then think. They often act and talk without sufficient forethought. They see themselves as "spontaneous."

Organization: High scorers are systematic and organized. They coordinate and plan all elements of a project. They thoroughly think things through.

Risk Taking: High scorers enjoy danger and risk. New activities and situation are not of concern. They are adventurous, unafraid, and fun loving. They are not necessarily impulsive about their activities; their action may be calculated and include a rational appreciation of the inherent danger.

Team Oriented: High scorers enjoy and believe in teamwork. They value team effort and team rewards. They do not enjoy working alone and may be inefficient when doing so.

Table 1. ALAPS Crew Interaction Scales Norms

Scale	Mean	*Sd*	Range
Deference	6.36	2.79	0-15
Dogmatism	5.90	3.08	0-16
Impulsivity	7.34	3.65	0-16
Organization	12.52	3.36	0-16
Risk Taking	12.17	2.96	0-16
Team Orientation	11.85	3.81	0-16

The air forces of the future will surely include more women and they will likely compete on an equal footing and may be represented in all cockpits. Statistical comparison of the men and women in the present sample on these scales resulted in a significant difference only on *Dogmatism* [male raw score equal to 6.06, female raw score equal to 4.64; $t = 4.9$ ($p < .00001$)].

DISCUSSION: The operators of the future will face an ever-changing enemy. As nation-states and political systems rise and fall so to will the nature of warfare and war machines. The cognitive abilities and personality make-up of combatants may need to change with both the enemy and technology. Pilotless aircraft and advanced spacecraft lend unique challenges to the psyche of the operator, as does rapid change from localized flare-ups to global nuclear threats. Experts in psychological research will be tasked to help aviators and policy makers keep the operator up with the rapid changes. As we invest increasingly large amounts of money into each individual airframe and mission, we must learn more about the human operator, whether that individual is a pilot or an operator in a virtual reality environment. ALAPS may aid selection of tomorrow's aviators, as it is an aviation-specific personality inventory. We plan to establish real-world, criterion validity by correlating findings on the ALAPS to behavioral measures, such as simulator flights, peer evaluations, and flight performance reports on mission-tested aviators.

Effective select-in measures using tools such as the ALAPS will be imperative with fewer cockpits, fewer aviators, and technologically sophisticated weapons systems.

Future research should consider cultural differences and expectations regarding the roles of men and women in a multinational collaborative effort. Crew coordination, squadron relationships, mission effectiveness, and flight safety may all be affected by the gender and cultural make-up of the unit.

REFERENCES:

1.) American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders, fourth edition.* Washington DC: American Psychiatric Press.

2.) Carretta TR, Retzlaff PD, Callister JD, & King RE. (In Press). A tale of two test batteries: A comparison of the Air Force Officer Qualifying Test and the Multidimensional Aptitude Battery. *Aviation, Space, and Environmental Medicine.*

3.) King RE & Flynn CF (1995). Defining and measuring the "right stuff": Neuropsychiatrically Enhanced Flight Screening (N-EFS). *Aviation, Space, and Environmental Medicine*; 66:951-956.

4.) Retzlaff PD, King RE, McGlohn SE, & Callister JD (1996). *The development of the Armstrong Laboratory Aviation Personality Survey (ALAPS).* AL/AO-TR-1996-0108.

Evaluating Concepts of Operation for Team / System Collaboration

Susan S. Kirschenbaum
Naval Undersea Warfare Center Division Newport
1176 Howell St.
Code 2211, Building 1171/1
Newport, RI 02841-5047
USA

1. SUMMARY

New systems demand new ways of working, both for the systems and for the users. This paper describes a methodology for evaluating the effectiveness and usability of complex, interconnected, collaborating systems. The methodology combines Exploratory Sequential Data Analysis, or ESDA (Ref 1), with a high fidelity, fully crewed, multiple scenario "Concept of Operation Exercise" that we at the Naval Undersea Warfare Center Division Newport call *COOPEX.* This paper describes the combination of ESDA and COOPEX methodologies in general and then reports on the results of one case study, the 1995 C3I[1] COOPEX for the New Attack Submarine

2. INTRODUCTION

New technology changes the way work is accomplished. Navy ships have always been composed of and controlled by many single-purpose pieces of equipment. The human was the integrating mechanism that glued these special-purpose system together. Collaboration between human and machine was largely a one way process in the same way that collaboration between the carpenter and the hammer is a one way process. The only general-purpose tools aboard Navy ships, planes, and almost any other "super-system," were writing implements. Paper, pencils, grease pencils and plexiglass were used for plotting, computing, recording, communicating, and facilitating collaboration.

Increasingly, modern computerized systems have the potential to be virtually any component. General-purpose hardware can perform many different tasks by the use of many different special purpose pieces of software. Such systems sense, record, analyze and exchange information; communicate with operators; and initiate actions electronically. They are no longer just obedient servants, but active participants. They demand new ways of collaborating among and between systems and users. These new operational concepts can radically alter the way work is accomplished and, thus, the relationships among crew members and between users and systems.

At the Naval Undersea Warfare Center Division Newport, we combine Exploratory Sequential Data Analysis, (ESDA) (Ref 1), with a high fidelity, fully crewed, multiple scenario Concept of Operation Exercise that we call *COOPEX.* The remainder of this paper first introduces the foundation components, ESDA and COOPEX, and then brings them together to report on the results of one case study, the 1995 C3I COOPEX conducted at the Naval Undersea Warfare Center Division (NUWC) located in Newport, RI, USA. I do not present detailed results from this COOPEX because they are not generalizable to other systems. Just as much of what has been learned about Human-Computer Interaction is task specific, much of what we learn from a COOPEX is system specific.

3. EXPLORATORY SEQUENTIAL DATA ANALYSIS: ESDA

Observational studies use a methodology that includes audio and video-taping followed by analysis of the transcriptions. The transcripts are encoded either by a generalized analysis procedure such as GOMS (Ref 2) or by a unique, task-specific encoding scheme (Ref 3). The history and theory of ESDA is reviewed in (Ref 4). The study reported here modifies and extends that tradition. For example, in the behavioral ESDA process, all observers typically employ the same coding schema, with the goal of agreement among coders for high inter-rate reliability. The observations are at the level of short durations and high frequency events and event sampling and statistical data analysis are employed. In contrast, interaction analysis within the social tradition of ESDA uses long duration, low frequency events observed in naturalistic, field settings and analyzed by intuitive, qualitative methods. Both of these traditions aim for inter rater reliability by having more than one observer code identical records according to an agreed upon schema.

The study reported here blends these two traditions, and adds a new element, multiple observers with different expertise, observing different part of a large and complex set of human-machine systems. Events are both observed concurrently, by a team of domain experts, and video-taped for later detailed analysis of key sections. Encoding reliability and validity is largely taken a face value because each of the observers has a unique expertise. It is confirmed only by repeated examples of the same behavioral patterns in different segments of the

[1] Command, Control, Communication, and Intelligence.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

Concept of Operation Experiment (COOPEX) and by partial analysis of selected sections of the taped record.

ESDA makes specific reference to *sequential* data but the same methods and tools can also be used to analyze concurrent data. concurrency can be thought of as just one end of the spectrum of possible sequences, ranging from completely overlapping to totally non-overlapping. With multiple players, actions and tasks range from independent to cooperative and from concurrent to sequential. Often team activity is composed of a dance of independent, concurrent activity and interactive, sequential activity (see Figure 1). Figure 1 is only an example. Actually this dance is more complex as information, including task coordination information, can flow by numerous routes such as voice, computer connections, face-to-face conferences, and telecommunication.) To analyze this dance requires multiple observers to record the actions of the multiple players while they are doing both independent and cooperative tasks. The study reported here reports the use of ESDA tools and methods in a multi-player, multi-observer experiment.

Of necessity, analysis-time to scenario-time (AT:ST) ratios are kept low because the results must be reported within weeks. However, the ST portion of the ratio can be quite long, amounting to weeks of actual run time. Post-run analysis is often aided by systematic queries to extract observations relevant to the specific experimental questions that motivated the COOPEX and guided selection of scenario events.

One example of a COOPEX will be reported below. It investigated issues such as operability, communications, equipment placement, and manning levels in a full Command Center composed of many interacting systems and manned by a full submarine crew. The system simulations ranged from cardboard prototypes to fully operational, advanced development modules (ADMs). Equipment ranged from existing special-purpose militarized consoles to commercial, off-the-shelf hardware. All consoles were networked to a simulation engine running Navy-standard models for both the ocean environment and the participating ships (US and others). The equipment was assembled in a laboratory and placed within the physical dimensions of the planned Command Center. Each console was located according to the current plans.

The full range of COOPEXes that we have conducted ranges from plywood and paper mock-ups of very early prototypes to a fully computerized, touch-screen ship control station mounted on a tilt-table.

4. CONCEPT OF OPERATIONS EXPERIMENT/EXERCISE: COOPEX

Over the course of more than a decade the Naval Undersea Warfare Center Division Newport has developed and refined a process to evaluate proposed concepts of operating with new equipment in specially designed Concept of Operation Exercises (COOPEXes). COOPEXes are extensive, often month-long studies that place prototype systems and knowledgeable crews in a variety of demanding scenarios. They are conducted to determine if proposed systems can operate within new operational concepts and without imposing unworkable requirements on the humans or if modifications need to be made.

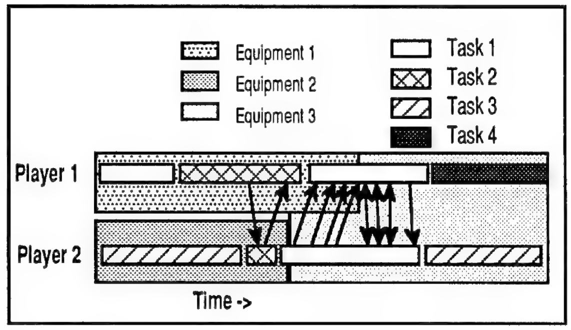


Figure 1: Concurrent and sequential team activity. Patterns indicate different tasks. Tasks are worked independently, although possibly concurrently, when unconnected. Arrows indicate synchronous interactions. Equipment may be used by a single operator or shared.

The "concept of operation" (COOP) is a proposed procedure and is central to defining military requirements for new systems or modifications to existing systems. It proposes a way of employing the new system for performing all of the anticipated functions. For example, a new, all electronic Ship Control System has been proposed for the next class of submarine. It will replace the largely mechanical systems on current submarines and will allow for a reduction from five to just two operators. Naturally, this new system will employ a very different COOP and thus the need for a COOPEX to test its ability to perform all the required functions.

COOPEXes are conducted in the laboratory, before committing the operational fleet to new equipment, operating procedures, or manning concepts. They always test the interactions of men[2] and machines and can involve single sub-systems, systems, or entire groups of linked systems. The objectives of specific COOPEXes can be very broad or narrowly defined. Either way, they determine the procedures used, define events to be combined into scenarios, and the variants to be examined.

Among the most important issues are manning and teaming concepts. Obviously, the number of trained operators required is a function of the likely missions and the kinds of tools available to carry out those missions. Aboard ship, the projected number of the crew impacts ship design in every area, from hotelling requirements (the number of berths and the size of the galley) to number of people available for damage control.

For the next generation of U.S. Navy attack submarine, COOPEXes have ranged from single systems (e.g., ship control) to the full, multi-system Command Center. The methodology requires domain experts to perform their

[2] At the current time all submariners are men so, for connivance, they will be referred to with masculine pronouns.

accustomed functions using prototype equipment. The equipment may range in fidelity from cardboard graphics to interactive, simulation-run, operational equipment. The first COOPEXes were conducted using rigidly controlled scripts because information was presented to crew members by static, pre-prepared 35mm slides projected onto screens. The screens were located where proposed displays were planned to be. Today, computer-based simulations allow the use of interactive advanced develop modules hosted on commercial, off-the-shelf (COTS) hardware. The spaces and layout of equipment are simulated in the laboratory, using proposed measurements and plywood housings to mimic militarized racks.

4.1 Why COOPEX?: Experimental questions

The specific objectives of each COOPEX differ, but the broad reasons for performing a COOPEX are similar. They are to test the proposed system or systems under as many varied and difficult scenarios as possible to assess their effectiveness for supporting the proposed operational tasks. The reason for maximally stressing the system is to find as many problems as possible, as early as possible. If a COOPEX is conducted while systems are still in the early design stages and the cost of change is minimal. Additional COOPEXes can be conducted to evaluate the effect of changes made.

Thus, the primary experimental question for all COOPEXes, as for many other HCI studies, is can the proposed system be used to satisfactory perform the required tasks.

4.1.1 Cost/benefit trade-off

Like life-saving surgery, a COOPEX is expensive to conduct, but far less expensive than the alternative. A COOPEX requires the full-time commitment of people and equipment for several weeks and the additional support of a smaller team in advance, for planning and afterward, for data analysis. However, the amount saved by conducting a COOPEX prior to committing to a major new system (like a new submarine design) far exceeds this initial expenditure, in both up-front costs and later revisions.

4.1.2 Typical questions

Typical issues investigated during a COOPEX can be categorized as human-human, human-machine, machine-machine, intra-system , and inter-system issues, where "system" is defined as the collection of people and machines working on some shared sub-component of the main task.

4.1.2.1 Human-human issues

Human-Human issues include crowding, face-to-face visibility, noise, communication, and command structure. These are typically social and organizational issues, but within the confines of a highly computerized set of tasks, a formalized social structure, and a confined space.

4.1.2.2 Human-computer issues

Human-machine issues range from ergonomic to operational questions typical of the domain of Human-Computer-Interactions and Human Factors. Thus common issues include anthropometry (seating), sensory-motor (visibility/legibility, screen layout, operation of Input-Output devices), and cognitive (location of functions, screen layout) functions. One of the most interesting questions that can be explored during a COOPEX is the use of multi-user equipment. An example of this issue will be discussed below, with respect to a horizontal, large-scale display (HLSD).

4.1.2.3 Computer-computer issues

Machine-machine issues include architecture, networking, data-base, and data-handling issues. These are not relevant to crew collaboration and will not be considered further.

4.1.2.4 Inter-system/Intra-system issues

Among the most interesting issues for any COOPEX are inter and intra-system questions. These include the layout of equipment, location of functions within equipment, and data needs by operators doing different component tasks.

Inter-system issues are almost impossible to assess any way other than a COOPEX because they involve the competing demands for (human, hardware, and facilities) resources and attention among systems. Thus they address questions such as can operators use communication equipment while performing their primary task? Can the network architecture support the data transfer demands of all systems without data loss or speed degradation? and Can operators access the information needed, do their own job, and pass the information to the next task in a timely manner and without impacting that ability of others to perform their tasks? This last is a question of both sequential allocation and adequacy of resources.

4.2 Participants

There are many people who make a COOPEX happen, each contributing their expertise. These include facilities, hardware, software, analysis, and operational experts. The participants are organized into teams with specific tasks, although there can be overlapping roles. All the teams are responsible to a single Production Manager.

I will focus here on two groups, the observation and analysis team and the operational crew. For greatest validity, the COOPEX crew should have considerable operational experience at the tasks that they will be doing. Moreover, they must be flexible enough to play-act where there are unrealities in the simulation and/or equipment due to the early-stage of the design process. Lastly, they must be imaginative enough to work within new procedures and operational concepts.

The observation and analysis team has a different role. In order to serve as knowledgeable observers, members must be thoroughly familiar with current operating practice, potential problems, and with the new, proposed, and simulated equipment. Lastly, they must be impartial, objective, and non-judgmental in their observations. While they ought to record *behavior*, they may want to give evaluations. As they are operational experts, their judgments are valuable and provision must be made to capture these expert judgments.

The team is lead by an expert in behavioral observation and ESDA techniques. The leader's responsibilities include planning, training observers, and analyzing the data. During

run-time she is trouble shooter, "gofor," substitute observer, and video recorder.

Phase	Activity	Participants
Preparation:	Simulation Equipment & space Scenarios	COOPEX teams: simulation facilities scenarios analysis
Training:	Train on equipment Practice scenarios	COOPEX teams: simulation analysis Participants: crew observers
Run-time:	Run scenarios Hot wash-up & questionnaries	Observers Participating Crew
Analysis:	ESDA analysis Questionnaire analysis	Analysis team

Figure 2: The phases of COOPEX production, activities, and players

4.3 Preparation Phase

The preparation phase begins with experimental questions, issues, and focus areas. This phase is often driven by concerns of potential users (the fleet) and of system developers and program managers. Questions often reflect design and acquisition choices. Typical questions are: Can this design or feature support the anticipated missions? How many operators/workstations are needed here? Which of these options is a better location for this equipment? What happens during this emergency? Will this information be visible from that location? What will the noise level be under these conditions?

These questions are essential to guide the COOPEX process. They dictate the required fidelity, data collection strategy, scenarios, and duration. However, specific issues are unique to each COOPEX and thus not relevant to this discussion. The remaining activities of the preparation phase; scenario, facilities, and simulation development, occur in parallel, coordinated by frequent meetings among the principals.

4.3.1 Scenario development and experimental design

Once the issues and experimental questions have been defined, scenarios are written that create opportunities to observe the key interactions. Scenarios are composed of plausible strings of events, including (1) normal, modal operations; (2) regular, difficult tasks; (3) infrequent, mission-critical tasks; and (4) catastrophic failures. Scenarios can range from fully scripted to improvisational. When a scenario is fully scripted, the players know what events will occur and even the displays are designated in the script. When they are more realistic, the events are defined only for the simulation and the participants know only as much as they would in the real world. How much the participants know about what will happen depends on the design stage of the systems. The earlier in the design process, the less likely it is that systems can support unscripted responses on the part of the operators.

To control for context effects, events are repeated in different orders in different scenarios. When time and resources allow, all scenarios are repeated with two or more crews. Critical events are always repeated. When more than one design option is being considered, relevant scenarios or events are repeated with each option.

4.3.1.1 Simulation, equipment, and environment

At any stage, a COOPEX must simulate portions of the world outside that impact the target system(s). These include components of targets (man-made and biological), off-platform messages, other parts of the ship, and the ocean (and/or land) itself. The choice of what aspects of the world to simulate depends on the planned interaction with the system and hypothesized effects. For example, when testing the use of touchscreen for ship control, a model of weather effects drove a motion platform action to create realistic test conditions. This degree of simulation fidelity was not deemed necessary to test C^3I systems.

As a COOPEX can take place early in the design process it may also need to simulate the working equipment and the facility where the equipment will be located. Thus, an early-stage COOPEX can included a simulated world and simulated prototype equipment in a simulated prototype room. In the past, simulation was crude and information changed by changing slides or flipchart pages. Today, the equipment interface is created by rapid prototyping and connected to a computer-simulated "real" world. Thus, operation can be relatively realistic, constrained only by the stage of the design process. Wherever possible actual, operational equipment is used. The addition of relevant physical characteristics of the environment, including space, noise, and lighting facilitates more accurate evaluation.

4.4 Training Phase

Training is one of the most important phases of any COOPEX. As the majority of the equipment is new, often, radically new, all role players and observers must be trained to operate the equipment. In addition to operation of the new systems, training must also include the new concepts of operation (strategies), changes in manning (who does what and who reports to whom), and training in experimental procedure.

The observers require addition training because they are typically domain experts, not experimenters. Many of the domain experts have served as instructors/observers in naval training exercises and schools. They are accustomed to reporting problems, but not to making non-judgment observations on behavior. Thus, observers must be trained in

behavior observation techniques. We use a special purpose program, MacShapa© (Ref 4), running on a Macintosh PowerBook™ to record time synchronized observations. Observers must be trained to use this equipment including the use of any preestablished coding scheme.

4.5 Data Analysis phase
Data analysis is a two step process. First, a "hot wash-up" is held immediately following a run. During the hot-wash-up role players review events with the observers. This retrospective gives observers a chance to ask questions and clarify the reasons for actions. Tapes can be reviewed for the role players, if desired. Role players also complete any questionnaires, including workload scales.

Post-experiment data analysis includes re-coding of comments, according to experimental questions; searching for critical incidents and key events to review in detail on tapes; counts of events, antecedents, and sequences; and comparisons across events, observers, equipment and layout variations, and role players. The time band for analysis events depends on the level of fidelity of the simulation and the questions being asked. It can range from seconds (e.g., a single operator's reaction time to a critical event) to tens of minutes (e.g., duration of events designed to evoke boredom) Sample results are provided below.

4.6 Run-time phase
During run-time a director serves as coordinator and time keeper. He or she starts the action so that simulation, observers, cameras, and role players are all synchronized. Events proceed according to the scripted scenario. Observers do have copies of the scenario so that they can be prepared for "emergencies" and know where to look for events.

Observers are assigned specific individuals or equipment and are briefed on key events and experimental questions for each scenario. Video cameras are placed to record all action and verbal communication.

5. CASE STUDY: C3I COOPEX

This study investigated the effectiveness of the entire Command, Control and Information (C^2I) system of a new class of U.S. attack submarine (NSSN). This system is currently being designed to meet anticipated needs well into the 21st century. A list of over 100 detailed issues/questions was generated by the designers and program managers. Major classes of issues included the ability of the crew to function on a less-paper ship, the noise levels and other factors associated with the plan to move sonar into the command center, use of several new pieces of equipment, manning requirements, new communication concepts, physical layout, and a new concept of operation for command itself.

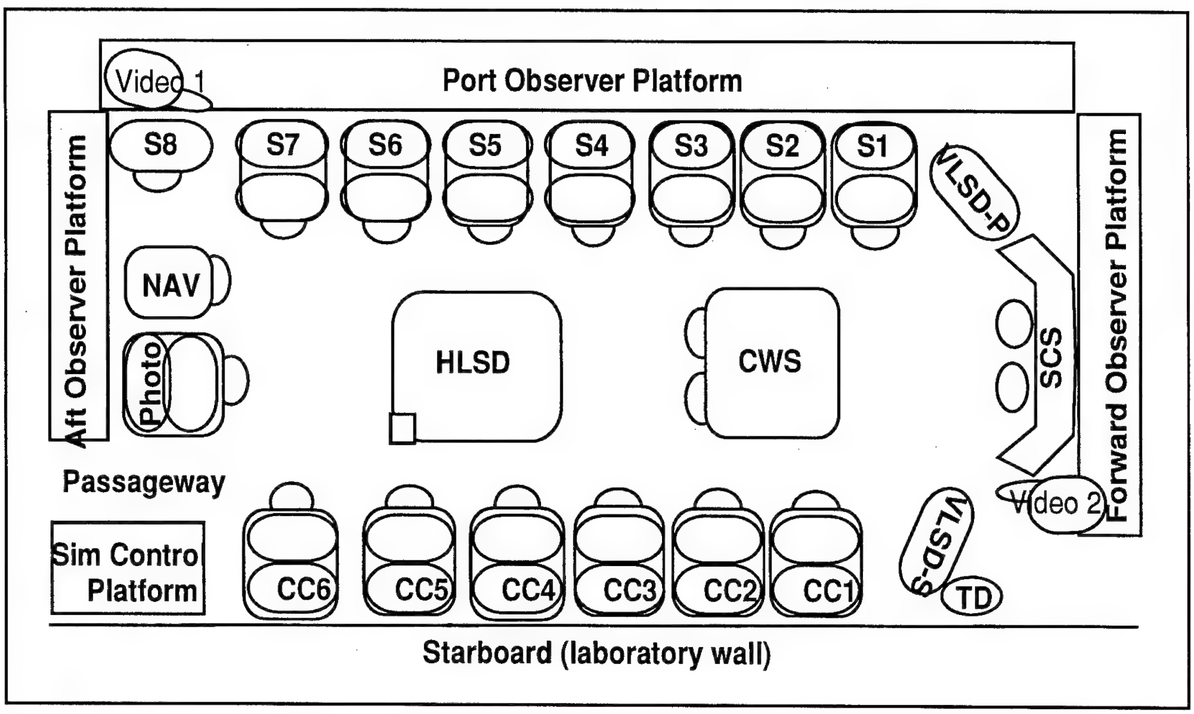


Figure 3: Workstation Layout

I shall concentrate here on one set of those study issues, the use of the horizontal large screen display (HLSD). The HLSD is a focus for collaboration and intended to support multiple roles. It replaces paper plots and is one of the keys to the reduced paper concept. The reader will also notice a strong relationship between the HLDS and the command workstation (CWS) located directly forward of the HLDS. The CSW, also new, is intended to support all command functions by bringing the information to the commander rather than requiring him to go to it. It was hypothesized that the proximity of the two items would facilitate efficient information gathering and usage.

5.1 Method

5.1.1 Facilities and Environment

The C^3I COOPEX took place in a specially constructed laboratory that was the same dimensions as the proposed Command and Communication Center (CACC). Overhead height was indicated by a string grid. Equipment was placed in the planned location. The location of other physical objects such as doorways, overhead hatches, and structural members were indicated.

A raised platform was build around three sides of the CACC was used for observers and camera placement. A rail was constructed at overhead height with movable platforms to support the laptop computers used by observers.

5.1.1.1 Plan of the Command Center

Figure 3 shows the workstation layout planned for the new CACC and the location of observation and simulation control platforms. The consoles are a mix of legacy and commercial, off the shelf hardware. In the figure the *Sonar* consoles, located on the port (left) side of the room, are denoted by "S." The *Combat Control* (CC) consoles are located on the starboard (right) side of the room. The "Photo" console is a special-purpose workstation designed to control the electronic imaging system that will replace current optical periscopes. These consoles (Sonar, CC, and Photo) contain two screens arranged one above the other. The CWS workstation has two adjacent screens, separated by a center section containing communications equipment. The two "VLSDs" are vertical, large-screen displays. The central HLSD is also a large-screen display, laid on its back so that the screen is horizontal. For COOPEX the large-screen displays were 24 inch monitors, but it is anticipated that they will actually be large flat panels. The two "SCS" displays for Ship Control and were simulated by cardboard pictures mounted on wooden screens. (An operational prototype was evaluated in a subsequent COOPEX.) Lastly, the "NAV" system was a commercial electronic navigation system using modern digital charts and GPS systems. When not in use for other purposes, the HLSD was used for navigation.

An observer platform served as the forward, aft, and port walls with the laboratory wall serving as the starboard wall. Observers were free to locate themselves where they could see best. Video cameras were trained to record all action. Video 1 recorded action in the forward and starboard portion of the CACC and Video 2 recorded action in the aft and port portion of the room. Due to the movement of crew and the early stage of some systems, detailed interactions with systems could not be reliably determined, but the observers and the video could capture what systems and which workstations were being used for what purpose.

Open microphones were placed above key equipment so that observers were able to hear discussions. Headset (point-to-point) communication was also monitored and recorded.

5.1.1.2 Computer Systems

A variety of special purpose and commercial, off-the-shelf (COTS) computer systems were used. One of the MMI objectives of this COOPEX was to test the ability of military and COTS systems to work together. The COTS systems were either Silicon Graphics machines or Hewlet Packert machines running under UNIX and X-Windows. All of the systems were networked so that they saw the same simulation databases. Scenarios were controlled and events synchronized by a single gameboard.

The observers employed MacShapa© (Ref 4) running on Macintosh Powerbooks™. A template was provided that contained two variables, a predicate "observations" variable and a text "comments" variable.

5.1.2 Scenarios

Scenarios had broad themes with specific events embedded within them. The role players were able to read abridged versions of the scenario scripts. these contained only the detail necessary for their roles. in most cases this mimicked the level of mission knowledge generally provided to a submarine crew, augmented by details necessary to compensate for non-functional system components.

5.1.2 Participants

5.1.2.1 Crew

The crew consisted of experienced current and retired submariners, supplemented by systems engineers when necessary. The core group were from a single ship and thus, accustomed to working together as a team. They included both officers and enlisted men, serving in the same role that they occupied aboard ship. Thus, the sonar operators and sonar Supervisor, the weapons operators and Officer, the Officer of the Deck (OOD), and others normally performed those functions at sea. Two of the participants were former submarine commanding officers (COs). They took the roles of CO and Fire Control Coordinator (FCC, the officer in charge of CC operators) The number of crew in the command center depended on the scenario and the events being simulated. It ranged from a low of three to a high of 35.

5.1.2.2 Observers

Four observers were trained to collect data during COOPEX. Each observer was assigned to observe a particular set of stations, locations, and personnel. The assignments were sonar (port) systems, combat control (starboard) systems, command workstation and ship control (center-forward) systems, and horizontal-large-screen display, navigation, and periscope (center-aft) systems. The center forward observer also tracked

movements of the senior officers who tend to rove. These officers are "anchored" at the CWS. In addition, the video operators observed, when possible. Observers were experts in their assigned area. The center and CC observers were former submarine COs or Executive Officers. The sonar observer was a former Sonar Supervisor.

5.1.2.3 Support personnel
Support personnel included facilities and computer system experts, simulation experts, and scenario experts. The scenario experts directed all activity, signaling the beginning and end of events to synchronize the simulation team and the observers.

5.2 Procedures

5.2.1 Training
Both crew members and observers were trained to operate the systems during the week prior to the start of the experiment. Observers were also trained to operate MacShapa and trained in behavioral observational techniques. Practice events were run also run to reinforce training.

5.2.2 Run-Time Procedure
The run-time procedure required a synchronized start for the four observers, two video-cameras, players, and all of the simulation components. To this end, the COOPEX director controlled the start and end of each event. During an event, timing was controlled by the simulation gameboards.

Each scenario took three to six hours. Each scenario was run twice, although sometimes, with some variation in the initial conditions; ocean conditions, equipment arrangement, or manning. Seven days, averaging 6 hours per day, were spent actually running scenarios. This produced 36 hours of tape from each of two cameras.

5.2.3 Post-run Follow-up
Two additional data collection procedures were used following each scenario, (1) immediate hot-wash-up and (2) next-day review of observations and tapes. These procedures were used to capture the expertise and in situ retrospective of the role players while the memory of events was fresh.

5.2.3.1 Hot wash-up.
Run-time recording was followed immediately by a "hot wash-up" in which players and observers discussed the scenario events. While not a full retrospective protocol, the hot wash-ups provided an opportunity for review of critical events by both role players and observers. In addition, it gave the observers an opportunity to ask for clarification or reasoning for observations that they had made.

Role players also completed subjective questionnaires to assess their judgments of such issues as workload, legibility and visibility of displays (including distant displays), the impact of communications on tasking, and associated issues. Questionnaires were correlated with key events and experimental questions addressed by the scenario. For example, both noise and crowding were addressed when the command center held the most people.

5.2.3.2 Next-day Review.
Two processes took place during the next-day reviews. The first employed a focus-group technique with analysis team, players, and interested representatives of the program managers. In these focus groups the participants used their collective experience to suggest resolutions to driving issues and to offer insight into problems that had not been anticipated.

The second technique involved reviewing the previous day's tapes for critical incidents. Tapes were reviewed with the players and with program managers' representatives so that the impact of these incidents could be assessed and corrective action proposed immediately. In some cases, this lead to equipment rearrangements that were tested on the next day.

5.3 Data preparation and analysis procedures

The compiled and recoded observations were used to select critical incidents for detailed analysis. Although frequently used in ESDA studies, sampling theory was not applicable because of the need to review *all* critical incidents, as judged by any of the observers.

All observations from a single scenario were combined into a single MacShapa spreadsheet. This facilitated coordinated analysis of data from multiple observers. By its nature, the data recorded in this COOPEX occurred at a rate of ~10 seconds to ~10 minutes, with a modal duration in the range of ~1 to 3 minutes. Precise timing was not required. The compiled spreadsheets did show considerable overlap among observers, indicating timing reliability.

The use of MacShapa spreadsheets facilitated locating key events for detailed analysis. In the structured, predicate variable, observers used predetermined key words. MacShapa's query language permitted search for additional relevant comments in the unformatted, free-text variable. For example, to locate observations of players using the HLSD, the query searched for "HLSD" or synonyms in the text. Selected cells were copied to a new predicate variable that combined relevant observations from all observers. The videotape of selected time periods could be reviewed for a careful analysis of critical events. Thus, location of these timestamped comments greatly reduced unproductive tape search and facilitated efficient location of all critical incidents and target events. In addition, MacShapa's various analysis and reporting capabilities allowed us to analyze concurrent events, antecedent events, and sequences of events.

Two techniques were used to locate critical sections of tape. One, as described above, was to use queries to select specific relevant observations. A second way to examine the many hours of data was to use the fast forward mode on the tapeplayer to find breakpoints, to follow player motions, and to find periods of intense usage for any workstation. This method was effective for some types of questions, especially those related to layout, ability of equipment to support multiple users, and player movements.

5.4 Results

The kinds of results possible with this evaluation method range from quantitative/descriptive (counts, durations, sequences,

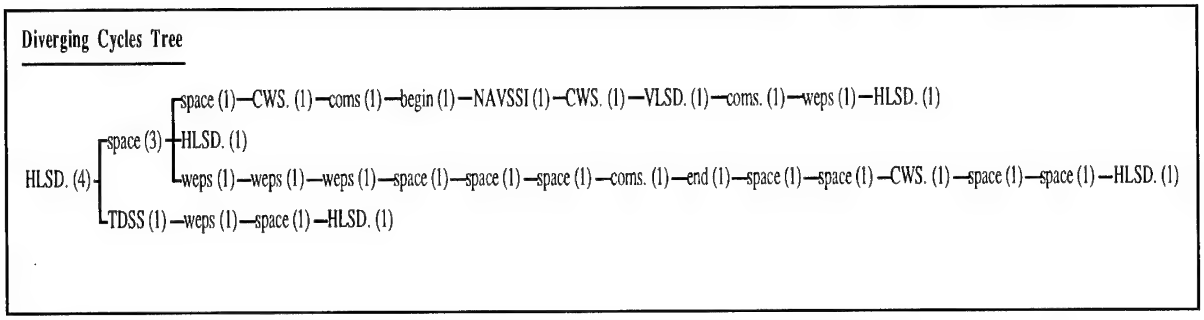


Figure 4: A trace of command movements around the command center during one scenario

distribution of events within scenarios, etc.) to qualitative (subjective judgments of participants and observers). However, these results do not lend themselves to analysis by inferential statistics. This, in part, is due to the exploratory nature of ESDA studies.

Figure 4 is an example of a cycle tree that traces all the observations between two HLSD records. This cycle tree comes from a single observer assigned to track the senior officers. This figure shows that here were four HLSD observations. Three of these were followed by space (arrangements) comments. Note that space was a frequent topic for observation. The top-most line records actions early in the scenario while the bottom two are later events.

The time element can be seen more clearly in Figure 5 (page 10-9), a timeline of all observations from the two central observers and from all observers combined (labeled "Panalysis:pred"). This shows the overlapping and sequential nature of observations.

The results showed heavy use of the horizontal large screen display, both for single users and for coordinating the tasks of several individuals. The largest number of individuals using the HLSD at any one time was nine. The largest bottlenecks were (1) lack of cursor control other than at the aft end of the console and (2) fixed display orientation. Figure 4 traces the activity of command between any two stops at the target console. Figure 5 gives timelines for console use as recorded from the point of view of three different observers. Differences are expected as each observer has a different assignment.

6. CONCLUSION

The method described here combines and extends the research method of exploratory sequential data analysis and the Navy method of concept of operation experiment to study the interactions between new systems and new concepts of operation.

The results range from descriptive (counts, durations, sequences, distribution of events within scenarios, etc.) to qualitative (subjective judgments of participants and observers). Implications from these results must be interpreted in terms of the trade-offs among system performance, operator training, and cost factors. These judgments are not within the purview of the researcher, but without the relevant data, responsible management decisions cannot proceed.

7. ACKNOWLEDGEMENT

This work was supported by the NSSN Program Office and the many participating sub-system Program Managers. The author also wishes to acknowlwedge the many COOPEX participants. This was the effort of a large group of dedicated women and men at NUWC, in Washington, and in the U.S. Navy Submarine Fleet. While there are too many to be named individually, special thanks goes to Scott McCarthy, ProductionTeam Director, and to the officers of Submarine Development Group 12.

8. REFERENCES

1. Sanderson PM, Fisher C, "Exploratory sequential data analysis: Foundations", Human-Computer Interaction, 9, 3&4, 1994, pp 251-317.
2. Card S. K., Morn T.P., Newell A., "The psychology of human-computer interaction", Hillsdale, NJ, Lawrence Erlbaum Associates, 1983.
3. Vortac OU, Edwards MB, Manning CA, "Sequences of Actions for Individual and Teams of Air Traffic Controllers", Human-Computer Interaction, 9, 3 &4, 1994, pp 319-343.
4. Sanderson PM, Scott JJP, Johnston T, Mainzer J, Watanabe LM, James JM, "MacSHAPA and the enterprise of exploratory sequential data analysis (ESDA)", International Journal of Human-Computer Studies, 41, 5, 1994, pp 633-681

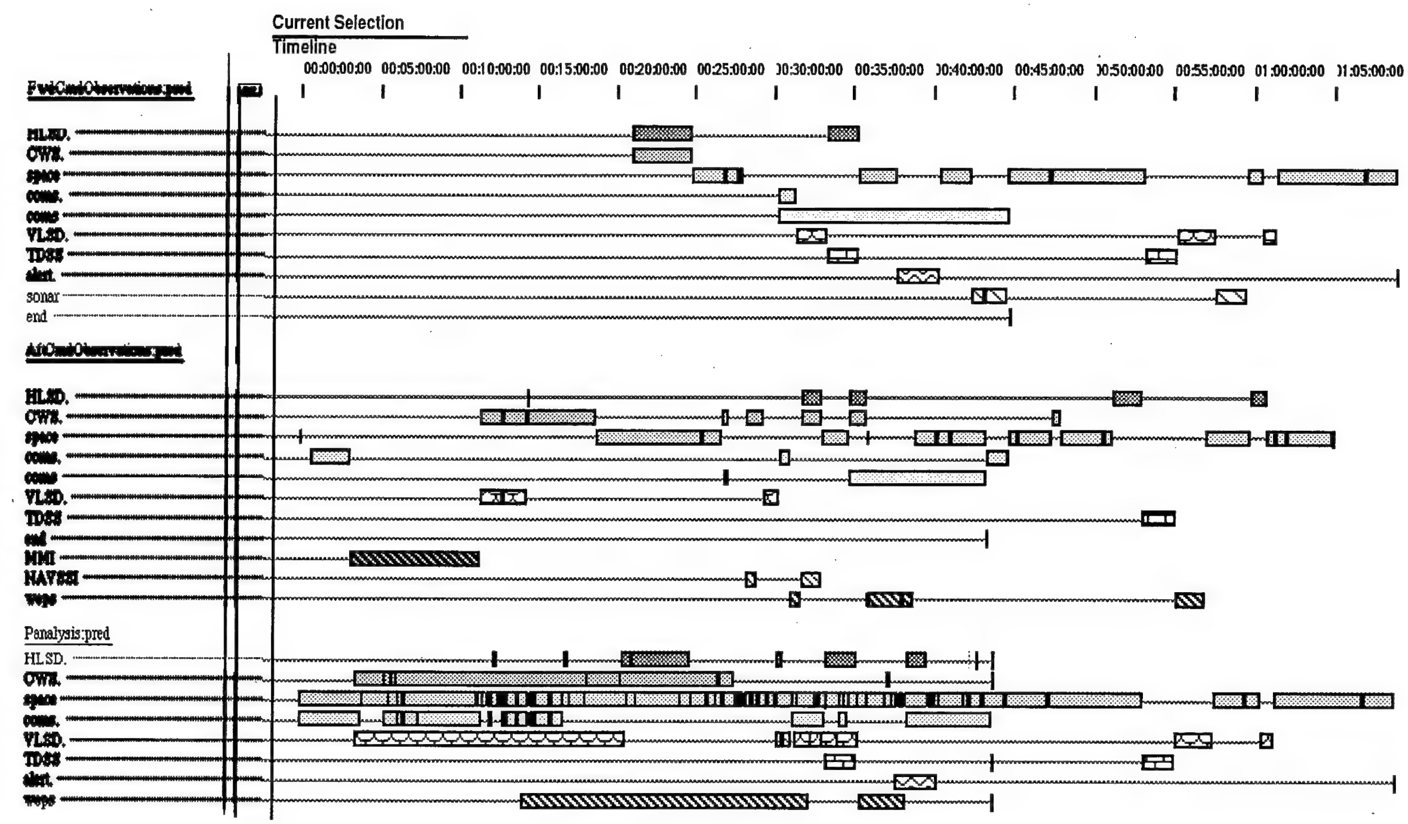


Figure 5: Time line from three observers of use of the horizontal large screen display. Note concurrent use of HLSD with other workstations and differences among observers.

A METHODOLOGY FOR MEASURING TEAM SITUATIONAL AWARENESS: SITUATIONAL AWARENESS LINKED INDICATORS ADAPTED TO NOVEL TASKS (SALIANT)

Elizabeth J. Muñiz
Renée J. Stout
Clint A. Bowers
Eduardo Salas
Naval Air Warfare Center Training Systems Division
12350 Research Parkway
Orlando, FL 32824-3275, USA
Attention: Code 4961

SUMMARY

Situational awareness has been recognized to be crucial for ensuring the effectiveness of teams performing in dynamic and complex environments. Given its criticality, researchers have called for reliable and valid measures of situational awareness that can be used as a basis for designing training (Ref 1; Ref 2). However, most of the available measurement techniques have been criticized as being insufficient for assessing situational awareness (Ref 3; Ref 4; Ref 5). Further, there is a dearth of research being conducted to measure team level situational awareness even though much of situational awareness is needed in team settings (Ref 2). Therefore, in this paper we describe a methodology for assessing team situation awareness. This methodology contains theoretically-based behavioral indicators of team situational awareness, which are adapted to specific task events. This methodology, termed Situational Awareness Linked Indicators Adapted to Novel Tasks (SALIANT), results in a behavioral checklist that can be used to behaviorally assess situational awareness in teams. A subsequent paper (i.e., Ref 6) describes empirical evidence testing the reliability and validity of this measurement approach.

INTRODUCTION

Today's military operators are required to perform in an environment that can be described as intense, dynamic, and abundant in information. Given the complexity of these operational surroundings, operators must work with one another to ensure that they obtain the information and resources necessary to accomplish their missions (Ref 7). Such interactions make operators dependent on one another to ensure the success of each mission. More specifically, these operators are required to function, as teams, in which team members assist one another by recognizing and attending to specific cues and events that could be vital to share during their missions. To identify such crucial information, these team members must rely on their situational awareness (Ref 2; Ref 8; Ref 1).

Situational awareness is vital for teams to perform effectively. Teams that do not rely on their situation awareness increase their chances of failure. For example, one often cited incident that illustrates the importance of situational awareness is the commercial aircraft that crashed into the South Florida Everglades in 1972 (Ref 9). More specifically, all crewmembers fixated their attention on a visual display suggesting a malfunction in the aircraft's landing gear, and they failed to notice that the auto-pilot became disengaged. This distraction prevented the crew from noticing a rapid and unexpected loss of altitude which led to ground impact. This accident is one of many incidents that have been related to breakdowns in situational awareness. In fact, analysis of mishaps caused by human related error report that a large percentage of these incidents identify situation awareness as a contributing factor (Ref 8; Ref 10).

Given the widely recognized criticality of situational awareness to the accomplishment of a variety of tasks, the concomitant need to assess reliably and validly situational awareness is of great importance. Without accurate measurement, effective training for enhancing situational awareness cannot be designed (Ref 1; Ref 2). Unfortunately, while there are variety of techniques that have been proposed to measure situational awareness, none of these have been explored fully in terms of their reliability and validity, and each has been criticized on a host of grounds (Ref 3; Ref 4; Ref 5). Moreover, there is a paucity of research, which has investigated measuring team level situation awareness even though much of situation awareness occurs in a team setting (Ref 2). Therefore, there is a dire need to develop measures

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

that attend to the team element in situation awareness. We contend that measures of team situational awareness must consider both cognitive and behavioral processes that indicate its presence or absence. The purpose of this paper is to describe a theoretically-based methodology for assessing team situational awareness. In the specific effort described here, we focused on behavioral processes related to team situational awareness. Before we elaborate on details about this methodology, we will briefly provide background information on situational awareness, situational awareness measures, and team situational awareness that contributed to the development of this methodology.

SITUATIONAL AWARENESS

As a first step in examining team situational awareness, it is imperative to understand the individual element of situational awareness. This level of research is crucial to acquire an understanding of a main component that influences team situational awareness: its team members. That is, the processes by which individuals acquire situational awareness can significantly impact the level of situational awareness acquired by a team.

Many definitions have been provided to explain components necessary for the achievement of individual situation awareness. A comprehensive review of these definitions is beyond the scope of this effort. (For an extensive review of situation awareness definitions please refer to Ref 11). The most commonly cited definition, however, is one provided by Endsley (Ref 12). She defined situation awareness as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and projection of their status in the near future" (p. 7).

Situation awareness components, such as the ones defined by Endsley (Ref 12) and other situation awareness researchers (Ref 11), provide an opportunity to develop measurement strategies to evaluate situation awareness in individuals. Similar to the conceptualization of situation awareness, there are a great number of measures that have been proposed to measure individual situational awareness. A review of these measures is also beyond the scope of this paper. However, we would like to note that many critics question the reliability and validity of these measures (Ref 3; Ref 4; Ref 5). In fact, such measures have been recommended to be used with caution (Ref 13; Ref 5). Further, most of these measures have been reported to be insufficient for capturing situational awareness at the team level (Ref 2; Ref 14). In the next section, we discuss the team component of situational awareness and its implications for measurement.

TEAM SITUATIONAL AWARENESS

Variations in team members' behaviors can influence the means by which they acquire situation awareness. That is, each individual is able to acquire a certain level of situational awareness based on his/her perceptions of cues in the environment and situations. Each team member's situational awareness is modified as he/she interacts with other team members who may have recognized or attended to an important piece of information that could be vital to update everyone's situational awareness. Hence, attaining situational awareness at the team level is more complicated given that additional team interactive processes play a significant role in the achievement and maintenance of situational awareness for <u>all</u> team members.

Over the years, several researchers have addressed the team element of situational awareness. More specifically, they have defined actions that suggest a team's level of situational awareness by reviewing literature on aviation teams, aviation mishap reports and/or aviators' responses to critical incident interviews. Several specific examples of behaviors that indicate high levels of team situational awareness have been provided in the literature, as well as those that suggest lower levels of team situational awareness. These are discussed in turn.

Behavioral Indicators of High Team Situational Awareness. Wagner and Simon (Ref 15) addressed the concept of team situational awareness in a training module. They defined team situational awareness as the crew's understanding of flight factors that can have an impact on the mission effectiveness and safety of the crew. The following flight factors were identified: mission objectives (e.g., flight plan, standard operating procedures); orientation in space (e.g., heading, altitude, airspeed of aircraft); equipment status (i.e., gauges, displays); external support (e.g., air traffic controller, navigational aids, ground guides); and personal capabilities (stress, fatigue, workload, skill). They suggested that aviators must recognize, process, and exchange information from several sources to develop and maintain situational awareness. While these flight factors are useful for determining what type of information should be exchanged for team situational awareness, there is no guidance as to how to exchange this information.

Bunecke, Povenmire, Rockway, and Patton (Ref 16) provided more specific behavioral recommendations as to how to exchange important information that would enhance the coordination skills of aviators which has some implications for team situational awareness. To begin, they suggested that crews should be kept advised of position specific information,

so that they are aware of the aircraft's 3-D position, and its relationship to any obstacle that might affect their ability to fly the aircraft. In addition, they noted that the team should confirm information when possible and challenge it, if in doubt, to resolve any incongruities that could affect their situational awareness. In addition, the team must communicate all pertinent information to increase the range of options and refrain from filtering or interpreting data that could be distorted in the process. Finally, the crew must discuss each other's human processing capabilities and plan on how the tasks would be shared during periods of high workload. Based on Bunecke et al.'s recommendations, team situational awareness can be captured by observing the crew's awareness of relevant information and discussions of how to share responsibilities during high workload periods.

Mosier and Chidester (Ref 17) focused on one of the components suggested by Bunecke et al. (Ref 16), namely, the exchange of information. Specifically, they investigated the relationship of information solicitation and transfer communications during simulated in-flight emergencies and abnormalities as indicators of situational awareness. Results indicated that the amount of information gathered by verbal exchanges had an impact on performance. Crews who performed better gathered new and relevant information before and after making crucial decisions. In addition, they re-checked old information (e.g., fuel status) to assure that the situation had not changed and that their assessment of it remained valid. Thus, Mosier and Chidester's findings suggest that team situational awareness involves the possession of information that is relevant for the crew to assess events occurring during their mission.

Foushee (Ref 18) noted that crews should use knowledge of changing situations and the immediate environment to enhance group process and aircrew performance. This knowledge would help the crew to identify potential problems and opportunities that can arise during flight. In addition, he suggested that crews should establish procedures to regulate processes, tasks, activities, and responsibilities, and take action to monitor the results of delegated assignments. Thus, other components that appear to be necessary for team situational awareness include awareness of environmental conditions, recognition of task problems and team members' actions.

Prince and Salas (Ref 19) provided a more extensive list of behaviors necessary for team situational awareness that was used to train and subjectively evaluate military pilots. These behaviors were extracted from literature reviews, mishap reports, and responses to critical incident interviews. They identified behaviors such as: identifying potential problems, recognizing the need for action, attempting to determine the cause of discrepant information, noting deviations, demonstrating on going awareness of mission status, and demonstrating awareness of task performance in self and others. Based on the results of this analysis, other behaviors that would manifest team situational awareness include the team being aware of their surroundings and each other's actions, and sharing relevant information with other team members.

Several years later, Prince (Ref 20) conducted additional critical incident interviews to obtain more detailed information of an aviator's perception on elements necessary to determine a team's level of situational awareness. For example, she asked more direct questions such *as, "When you are observing a crew, what do you look for as indicators of the crew's situation awareness?".* The responses obtained included actions such as being ready to answer questions, having contingency plans, briefing status frequently, responding quickly to radio messages, and considering passengers.

Behavioral Indicators of Low Team Situational Awareness. Prince (Ref 20) also identified behaviors indicating poor team situational awareness such as lack of communication, lack of listening, having an argumentative crew, not noticing mistakes, overloaded crew members, and being unaware of problem consequences. Similar behaviors were delineated by Leedom (Ref 21) who extracted this information from Army aviation accident reports. He noted that most of these incidents were caused by failing to inform team members of actions taken and not acknowledging communications or resolving conflicts.

Schwartz, (Ref 14) to instruct aviators, has used examples of behaviors indicating team situational awareness loss. This training module provides a more extensive list of poor indicators of team situational awareness. These behaviors included: (a) ambiguity of information when two sources do not agree; (b) fixation or preoccupation on one item or event and excluding all other information; (c) confusion, sense of uncertainty, and anxiety about a particular situation; (d) no one flying the aircraft or monitoring the flight progress; (e) crew not performing visual lookout procedures (i.e., not looking out the window); (f) deviating from standard operating procedures; (g) using undocumented procedures that are not prescribed in flight manuals; (h) violating limitations or minimum operating standards; (i) failure to resolve conflicts and/or discrepancies; (j) failure to meet targets (e.g.,

altitude, headings); and (k) incomplete communications resulted from withheld information.

Schwartz (Ref 14) suggested that the aforementioned actions were identified as cues of an "error chain" in progress. He suggested that a crew's error is a result of a chain of events, which often leads to mishaps. A "chain of errors" can be sequential, related or unrelated to each other, and in many occasions not readily salient to the crew. Thus, a single overpowering factor would rarely be the reason for an accident. That is, the mere presence of a single behavior should not be concluded that the crew has completely lost situational awareness. Judgment and discretion are required to determine whether an "error chain" is in progress. Schwartz's "error chain" argument implies that a single behavior cannot be considered the sole indicator of team situational awareness. Rather, behavioral patterns or sequences are more likely to indicate the level of situational awareness in teams.

In summary, the available literature theorizes on several team behaviors associated with team situational awareness. The combination of behavioral patterns in a particular team context can be used to determine a team's level of situational awareness (Ref 14). This would suggest that some component of team situational awareness can be observed behaviorally, which has some implications for measuring team situational awareness. In the next section, we describe a methodology for assessing behaviorally situational awareness in a team context.

TEAM SA MEASUREMENT METHODOLOGY: SALIANT

The methodology described here contains categories of situational awareness indicators that were derived from the literature and events that are embedded in scenarios to elicit each of these categories of required situational awareness behaviors. Hence, the behavioral indicators of team situational awareness are theoretically-linked to situational awareness and tied to or adapted to, specific task scenarios. This approach is termed Situational Awareness Linked Indicators Adapted to Novel Tasks (SALIANT). Using this methodology, within a specific application, results in a behavioral checklist that can be used to assess the patterns of behaviors exhibited by a given team to determine their level of situational awareness. The development of SALIANT is comprised of five phases. Each phase is crucial to ensure that the construct of team situational awareness is measured consistently and objectively. What follows is a brief description of each phase taken to develop SALIANT.

PHASE I. ***Delineation of Behaviors Theoretically Linked to Team Situational Awareness.*** The behaviors reviewed in the previous section that were theorized to be associated with team situational awareness are delineated in Table 1. In turn, these behaviors were clustered into five categories based on common elements. These categories included: (a) demonstrating awareness of surrounding environment; (b) recognizing problems; (c) anticipating a need for action; (d) demonstrating knowledge of tasks; and (e) demonstrating awareness of important information.

Table 1. Behavioral indicators of team situational awareness.

Demonstrated Awareness of Surrounding Environment
• Monitored environment for changes, trends, abnormal conditions (Ref 20)
• Demonstrated awareness of where he/she was (Ref 16)
Recognized Problems
• Reported problems (Ref 19; Ref 18)
• Located potential sources of problem (Ref 19)
• Demonstrated knowledge of problem consequences (Ref 20)
• Resolved discrepancies (Ref 14)
• Noted deviations (Ref 19)
Anticipated a Need for Action
• Recognized a need for action (Ref 20; Ref 19; Ref 18)
• Anticipated consequences of actions and decisions (Ref 20)
• Informed others of actions taken (Ref 21)
• Monitored actions (self & others) (Ref 19)
Demonstrated Knowledge of Tasks
• Demonstrated knowledge of tasks (Ref 14)
• Exhibited skilled time sharing attention among tasks (Ref 14)
• Monitored workload (self & others) (Ref 20)
• Shared workload within station (Ref 16)
• Answered questions promptly (Ref 20)
Demonstrated Awareness of Information
• Communicated important information (Ref 16)
• Confirmed information when possible (Ref 16; Ref 20)
• Challenged information when doubtful (Ref 19; Ref 21; Ref 16)
• Re-checked old information (Ref 17)
• Provided information in advance (Ref 19; Ref 14)
• Obtained information of what is happening (Ref 18)
• Demonstrated understanding of complex relationships (Ref 16; Ref 14)
• Briefed status frequently (Ref 20; Ref 14)

PHASE II. ***Development of Scenario Events.*** The development of scenario events was based on two considerations: (a) the opinion of a subject matter expert to ensure the operational relevancy, and (b) a team task analysis to ensure the scenario is complex, dynamic, and requires team member interdependency (Ref 22).

A brief example of a recently developed scenario required teams to complete an exercise that simulated a reconnaissance and target destruction mission. In this scenario, the team

was provided with a pre-planned route that contained an ingress route, a target destruction area, and an egress route. The team was required to fly from one point to another until they arrived to a target area. While the team followed the ingress route, they experienced a series of events. For example, their radio frequency was shared with other aviation units. In addition, one aviation unit reported being too close to the team which required a deviation from their pre-planned route.

Once the team arrived to the target area, the team was required to destroy all ground targets in sight, and then return to base by following the egress route. During their egress route, the team experienced radio problems with the radio tower, and it was required to take action (i.e., increase their altitude). After the team re-established contact, one aviation unit interacted with the team by asking them whether they found the targets, and how many targets were destroyed. Finally, when the team flew back to their landing site, they shared their radio frequency with another aviation unit who experienced a cargo swift shift swing.

The scenario just described contains specific events designed to prompt teams to manifest team situational awareness behaviors. More specifically, each event or task was tied to one of the behavioral indicators of team situational awareness (identified in Table 1). The requirement of linking behaviors to scenario events led to its naming SALIANT. Once the scenario is developed, we can continue with the next three steps.

PHASE III. ***Identification of Specific, Observable Responses.*** The scenario described in Phase II was developed to prompt teams to measure situational awareness in an aviation team context. Hence, we transformed the behavioral indicators into more specific observable responses based on the five flight factors identified as crucial for attaining crew situational awareness (Ref 15). As mentioned earlier, Wagner and Simon suggested that factors such as mission objectives, orientation in space, external support, equipment status, and personal capabilities are necessary for a crew to maintain and support team situational awareness. In addition, in each of these five categories, we included specific behaviors that take in consideration the anticipatory component necessary to acquire high levels of situational awareness (Ref 2; Ref 8; Ref 11; Ref 23). For example, for the event in which the team experienced radio problems with the Air Mission Commander (AMC), the team was required to climb to a higher altitude to re-establish contact. In this event, we attempted to measure several behaviors, one of which was "informed others of actions taken." This "generic" behavior was translated into the specific behavior: "Team member mentioned that he would be climbing to a higher altitude before being asked by the other team member." Figure 1 illustrates an example of a matrix that contains several specific responses for the scenario described in phase II.

Figure 1. Matrix Depicting Specific, Observable Responses

Generic Behaviors	Specific Behaviors				
	Mission Objectives	Orientation in Space	External Support	Equipment Status	Personal Capabili-ties
(1.1) Reported Problems				Team member verbalized they have lost radio contact with AMC	
(1.2) Resolved Discrepancies				Team member suggested to climb to a higher altitude to reestablish contact	
(1.3) Informed Others of Actions Taken				Team member mentioned that he will be climbing to a higher altitude before being asked by the other team member	
(1.4) Noted deviations		Team member called out heading deviations before being asked by the other team member			

PHASE IV. ***Development of a Script.*** A script was developed to ensure consistency across teams. The script contained specific instructions as to

when to introduce each event, specific information that will be provided to the teams, and how to respond to the teams.

PHASE V. ***Development of an Observation Form.*** A structured observation form was developed to rate the specific observable behaviors identified in phase III (see figure 2). The form contains four columns. The first column was used to identify the scenario segment of a specific scenario. The second column contains the events that were introduced in the scenario. The third column included specific observable responses to these events. Next to this column, we included a code number that links each specific response to one of the 24 behaviors identified in Phase I and one of the five categories suggested by Wagner and Simon (Ref 15). The last column was developed for the observer to check off the presence of this behavior.

Figure 2. Example of SALIANT observation form

Scenario Segment	Event	Acceptable Response	Code	Hit
From Point Golf to India	Team lost radio contact with AMC	Team member discussed possibility of radio contact loss.	EQ 1.1	X
		Team member suggested to climb to a higher altitude to re-establish contact.	EQ 1.2	X
		Team member mentioned that he will be climbing to a higher altitude.	EQ 1.3	

In general, SALIANT methodology appears to be promising to study team situational awareness. The main reason is that this measurement methodology is theoretically based, which ensures we are assessing actual manifestations of team situational awareness. In addition, SALIANT allows the opportunity to evaluate objectively team situational awareness, which maximizes its reliability. That is, consistency among raters should be obtained by delineating in advance the expected responses of a team to each event. A similar approach was used by Fowlkes, Lane, Salas, Franz and Oser (Ref 24) to assess aircrew coordination behaviors and found high levels of agreement among raters. Further, the development of a scenario that is operationally relevant to the team maximizes the opportunities to observe team behaviors that can actually occur in real life situations. Finally, while this effort was described to be used in an aviation environment, we believe that this methodology can be used in other team settings.

Although this methodology appears to be beneficial to study team situational awareness, we have to acknowledge that there are some limitations in using SALIANT. First, while we recognize the importance of cognitive processes involved in team situational awareness, this methodology does not assess any of the cognitive components of team situational awareness. Second, the behaviors used as a basis for this methodology only take in consideration those that manifest high levels of team situational awareness. A similar strategy needs to be developed that incorporates indicators of poor situational awareness. Third, the development of an operationally relevant scenario requires extensive knowledge about the tasks, standard operating procedures, and regulations required for a team to perform their missions. Finally, the delineation of the most effective responses is a labor intensive effort. In fact, in some occasions, some of the expected responses may not occur naturally, or other responses that were not identified can be exhibited.

To offset these limitations, we recommend continuing research with SALIANT. Specifically, we recommend that this methodology be tested to determine its construct validity. For example, constructs theorized by the literature as moderators of team situational awareness (e.g., planning, Ref 25; communication, Ref 26; shared mental models, Ref 2) should be examined to determine their relationship to SALIANT. In addition, this methodology should be evaluated with teams that perform in real operational environments. The next paper by Bowers, Barnett, Weaver, and Stout (Ref 6) describes an empirical study designed to assess the utility and validity of the SALIANT measurement approach for tactical aviation.

REFERENCES

1. Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. Human Factors 37, (1), 123-136.

2. Stout, R. J., Cannon-Bowers, J. A., & Salas, E. (1996). The role of shared mental models in developing team situational awareness: Implications for training. Training Research Journal, 2, 85-116.

3. Cooke, N. J., Stout R. J., & Salas, E. (in prep). A knowledge-based approach to measuring team situation awareness.

4. Fracker, M. L. (1991a). Measures of situation awareness: Review and future directions (Rep. No. AL-TR-1991-0128). Wright- Patterson Air Force Base, Ohio: Armstrong Laboratories.

5. Fracker, M. L. (1991b). Measures of situation awareness: An experimental evaluation (Rep. No. AL-TR-1991-

0127). Wright-Patterson Air Force Base, Ohio: Armstrong Laboratories.

6. Bowers, C. A., Barnett, J., Weaver, J., & Stout, R. J. (1998). Empirical validation of the SALIANT methodology. The 1st Annual Symposium/Business Meeting of the Human Factors & Medicine Panel on Collaborative Crew Performance in Complex Systems, Edinburgh, United Kingdom.

7. Salas, E., Dickinson, T. L., Converse, S., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.) Teams: Their training and performance (pp. 3-29). Norwood, NJ: Ablex.

8. Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. In R. D. Gilson (Ed.), Human Factors. Special Issue: Situation Awareness, 37 (1), 65-84.

9. National Transportation Safety Board (1973). Aircraft accident report: Eastern Airlines 401/L-1011, Miami, Florida, December 29, 1972 (NTSB/AAR-73-14). Washington, DC: Author.

10. Hartel, C. E. J., Smith, K., & Prince, C. (1991). Defining aircrew coordination: Searching for mishaps for meaning. Paper presented at the 1991 sixth International Symposium on Aviation Psychology, Columbus, OH.

11. Dominguez, C. (1994). Can SA be defined? In M. Vidulich, C. Dominguez, E. Vogel, & G. Macmillan (Eds.). Situational awareness: Papers and annotated bibliography (Report AL/CF-TR-1994-0085, pp. 5-15). Wright-Patterson Air Force Base, OH: Air Force Systems Command.

12. Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. Proceedings of the Human Factors Society 32nd Annual Meeting, 97-101.

13. Shrestha, L. B., Prince, C., Baker, D. P. & Salas, E. (1995). Understanding situation awareness: Concepts, methods, and training. Human Technology Interaction in Complex Systems, 7, 45-83.

14. Schwartz, D. (1990). Training for situational awareness. Houston, TX: Flight Safety International.

15. Wagner, M. & Simon, R. (1990, August). Development of measures of crew coordination. (Contract DAHC35-89-D-0030).

16. Bunecke, J. L., Povenmire, H .K., Rockway, M. R. & Patton, M. W. (1990). The awareness component of cockpit resource management. Proceedings of Psychology in the Department of Defense, (pp. 259-263). USAF Academy Department of Behavioral Sciences and Leadership.

17. Mosier, K. L., & Chidester, T. R. (1991). Situation assessment and situation awareness in a team setting. Proceedings of the 11th Congress of the International Ergonomics Association, Paris.

18. Foushee, H. C. (1984). Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. American Psychologist, 39 (8), 885-893.

19. Prince, C., & Salas, E. (1993). Training and research for teamwork in the military aircrew. In E. L. Wiener, B. G. Kanki & R. L. Helmreich (Eds.), Cockpit resource management. Orlando, FL: Academic Press.

20. Prince, C. (1998). Guidelines for situation awareness training. (Preliminary rep by NAWCTSD, UCF & FAA). Orlando, FL.

21. Leedom, D. K. (1990). Aircrew coordination training- A new approach. Fort Rucker, Alabama: U.S. Army Research Institute Aviation R&D activity.

22. Salas, E. & Cannon-Bowers, J. A. (1997). Methods, tools, and strategies for team training. In M. A. Quiñones & A. Dote (Ed.), Training for 21st century technology: Applications for psychological research. Washington, D.C.: APA Press.

23. Vidulich, M. (1994). Cognitive and performance components of situation awareness: SAINT Team Task One report. In M. Vidulich, C. Dominguez, & G. Mc Millan (Eds.), Situation awareness: Papers and annotated bibliography (U) (Rep. No. AL/CF-TR-1994-0085). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

24. Fowlkes, J. E., Lane, N. E., Salas, E., Franz, T., & Oser, R. (1994). Improving the measurement of team performance: The TARGETs methodology. Military Psychology, 6, 47-61.

25. Stout, R. J. (1994). Effects of planning on the shared understanding of team member information requirements and efficient communication strategies. Unpublished doctoral dissertation, University of Central Florida, Orlando.

26. Bowers, C. A., Baker, D. P., & Salas, E. (1994). Measuring the importance of team work: The reliability and validity of job/task analysis indices for team training design. Military Psychology, 6, 205-214.

Empirical Validation of the SALIANT Methodology

C. Bowers
J. Weaver
J. Barnett
University of Central Florida
4000 Central Florida Blvd
Orlando, Florida 32816-1390, USA
R. Stout
Naval Air Warfare Center Training Systems Division
Orlando, Florida 32826-3224, USA

SUMMARY

Past research has indicated the importance of considering the situation awareness (SA) construct as it might apply to team performance. This report attempts to contribute to our understanding of this research area, through the development of a measure of team situation awareness. The method for the development and preliminary validation of the measure is described herein (i.e., SALIANT; Situation Awareness Linked Instances Adapted to Novel Tasks). The methodology was developed to be one that would be appropriate for variety of team applications. However, this report describes our preliminary validation of the measure with aircrews. Specifically, the effort sought to assess whether the measure would demonstrate expected associations with performance indices and to compare its utility to an existing SA methodology (i.e., SAGAT). The report describes our findings regarding the effectiveness and benefits of the measure as well as providing recommendations and caution for its future use.

1 INTRODUCTION

Many present day occupations require the effective performance of teams. However, our understanding of the factors that ensure such effective performance is still lacking. One such factor that has been identified as critical for such task performing teams is situation awareness. In fact, it has been argued that situation awareness is particularly important for tasks that are complex, knowledge rich, and dynamic in nature (Ref. 1). Factors such as these characterize many team task performance situations (Ref. 2). Examples of instances where loss of situation awareness is often disastrous abound including flight operations of air crews, command and control operations, and the performance of medical teams. All of these share the requirement for an accurate awareness of the situation, in order to effectively accomplish goals. Situation awareness has been defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Ref. 3, p. 36).

Unfortunately for team researchers and practitioners, much of the research conducted to date has attempted to improve our understanding of the situation awareness construct as it applies to individual task performance situations, while contributing little to our understanding of the manner in which situational awareness might apply to team task performance situations. This issue is becoming increasingly important because of the nature of many team task performance situations, when success or failure may be related to the team's overall awareness and assessment of the situation and its consequent actions.

An important first step in learning about team SA is to develop appropriate measurement instruments. Such measurement instruments would allow researchers to distinguish between teams with good SA versus those with poor SA. Such measures would then be of use to assess the extent that level of SA is predictive of and related to performance indices. Therefore, the current project sought to develop a team situation awareness methodology that could be modified for teams in different domains. However, the current project sought to develop the measure and establish preliminary validation of the measure via application to aircrews.

Salas, Prince, Baker, and Shrestha (Ref. 4) have argued that team SA is much more complex than simply the combination of SA of the individual members. For example, team SA must necessarily also include assessment of activities unique to team task performance situations such as information sharing and coordination (Ref. 4). In considering past efforts to define individual SA and the few efforts that have been made to begin to define team SA, Salas and his colleagues offer the following proposition. Specifically, they suggest that team SA is composed of two components: individual SA and team processes such as teamwork behaviors. These authors offer two recommendations for research to improve our understanding of team SA. First, they argue that team SA research should use past teamwork research as a foundation from which to select teamwork behaviors to be considered, and second, they argue that our understanding of individual SA, and thus team SA, should be extended by placing emphasis on the dynamic nature of the situation.

Consequently, building upon these recommendations, direct implications for measurement can be gleaned. In short, Salas and his colleagues argue that the use of embedded events within scenarios have already been used successfully in past team research as a technique for eliciting behaviors of interest. Furthermore, efforts at measuring team SA should adopt methods that enhance rater reliability and validity. Finally, these authors cite the TARGETS (Targeted Acceptable Responses to Generated Events or Tasks; Ref. 5) methodology as an exemplar of an assessment technique with these qualifications.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

Fowlkes and her colleagues (Ref. 5) developed the TARGETS methodology for aircrew coordination training. This structured observation methodology utilizes carefully structured scenarios to provide the opportunity for teams to demonstrate skills. The responses deemed to be acceptable at given points within the scenario are determined apriori by SMEs (subject matter experts). This allows rates to then assess whether the acceptable response is present or absent at the appropriate portions of the scenario.
In summary, the current study will build upon the work of earlier researchers by determining the extent that this methodology might be appropriate and useful as a method of assessment of team SA.

Development of the measure. The measure uses a simulation or role-playing exercise as a vehicle to elicit the behaviors. Subject matter experts (SME's) develop a scenario representative of the domain in which the teams are to be measured; surgery or trauma rooms for medical teams, a command center exercise for military teams, aircraft simulators for aviation teams, etc. In this scenario, teams with good (or poor) situation awareness could be expected to show certain behaviors at particular points, or in response to certain situations. The next step is for scenario developers to list specific behaviors that would be expected from teams with good vs. poor awareness of the situation in question.

Once the scenario is developed, SMEs then predict what behaviors would indicate the team's level of SA at different points in the scenario. A score sheet is developed listing these expected behaviors. These expected behaviors should then be categorized as indicating good SA or poor SA. Teams are then placed in the scenario and their actions recorded, typically by video or audio tape. Later, raters can review the recordings and score the team's performance by marking each expected behavior on the score sheet. The total number of positive and negative points is summed, giving a relative score for the team.

Because the precise behaviors that indicate SA levels are normally situation specific, it is difficult to provide exact guidelines for choosing expected behaviors. However, other investigators (Refs. 6, 7, 8, and 9) have illustrated behaviors that indicate levels of situation awareness. These behaviors can form the foundation for selecting more scenario specific behaviors and were utilized to develop the measure reported here.

Although the behaviors described by previous research form a valuable basis for developing a scenario specific measure, there is considerable overlap among the behaviors listed, and many of the described behaviors are scenario, or at least domain, specific. In our study, we simplified the task of applying behaviors from previous research by analyzing the literature and grouping listed behaviors into categories. While there could be alternative ways of grouping the behaviors, we adopted four categories: orientation, communication/information management, workload management, and problem resolution because these areas appear to encompass the behaviors that indicate levels of situation awareness for both individuals and teams.

Orientation refers to an awareness of where the team is in space, if spatial awareness is relevant; or, if not, the team's awareness of where they are in relation to accomplishing their mission. For example, for flight crews, an awareness of their position in physical space is important, but their position in space is correlated with their place in the mission. For example, if you are only a few miles from your final destination, you are normally near the end of the mission. However, for teams that are stationary, like medical trauma teams, spatial position is irrelevant. Yet, a trauma team must be aware of how they stand in their mission of stabilizing someone who is critically injured. For them, it is the relative nearness to their goal that is important.

Communication is vitally important for maintaining situation awareness within a team. Also important is how well the team manages information. These two constructs are related and thus, are counted together. Behavioral indices of problem resolution were included in light of past research (Ref. 8) that has indicated that the identification and reporting of potential problems is related to aviation team situation awareness. Similarly, research conducted by Bunecke and colleagues (Ref. 6) has recommended that a team's situation awareness is facilitated by its ability to share and monitor workload in the cockpit.

In summary, the current effort developed the SALIANT measure of team situation awareness that assesses a subset of the behaviors expected when teams are in a state of good situation awareness. The relevant situation awareness behaviors were identified via literature reviews and scenario developers created a flight task suitable for use by trained pilots in which these behaviors were "built in." This simulation acted as the vehicle to test the elicitation of these behaviors. Behaviors were then rated using the measure. Finally, the data yielded by these ratings were then utilized to test the extent that they would be, 1) significantly related to indices of performance, and 2) related to an existing method of assessing SA (i.e., SAGAT) and yet contributing unique variance to indices of performance.

2 METHOD

Participants

Participants were 30 pilots and flight instructors from the Comair Aviation Academy in Sanford, Florida. Their experience ranged from 140 to 2700 flight hours in propeller-driven aircraft. Participants indicated they had a fair amount of experience flying with other crewmembers in instructor-student relationships, but little experience flying with other pilots of similar experience.

Apparatus

Flight simulation. A simplified flight simulator was constructed using an IBM-style personal computer (PC) with commercially available flight simulation software. The PC was a high speed (166Mhz) Pentium™ with graphics, video, and sound cards optimized for simulation software. It was outfitted with a joystick that included an integral throttle wheel, and rudder pedals.

The software chosen simulated a US Army *Apache* attack helicopter, which has a crew of two; a pilot and a copilot/gunner. The control configuration allowed the joystick to be used as a cyclic control, the throttle wheel as the collective, and the rudder pedals as the helicopter's tail rotor control. A "cockpit" was constructed consisting of two computer monitors connected to the central processing unit via a video splitter. The monitors were side-by-side on a table with a screen between them to prevent crewmembers from observing each other's actions. The screen forced the participants to verbalize questions, rather than observe the other team members' actions. The "pilot's" side had the joystick and rudder pedals, while the "copilot/gunner's" side contained the computer keyboard. Pilot studies showed the pilot did not need the keyboard to fly the simulator, while the copilot/gunner used it exclusively to operate the sensors and weapons. The experimenter sat behind a screen near the participants and controlled the experiment, while also playing the roles of outside agencies who were required to communicate with the team.

The simulator included an intercommunications system (intercom) consisting of an amplifier and lightweight headsets with microphones for both participants and the experimenter. The system was used in the "hot mic" mode, where the microphones were always active and the participants need not press a switch to talk.

Simulator training. A videotape was locally produced to train the participants regarding how to fly the simulator and use the sensors and weapons. This ensured that each team of participants would receive the same basic information. The training portion of the videotape lasted about 15 minutes. Following the training portion, a mission briefing section introduced the participants to the experimental mission and objectives. This section lasted about 12 minutes.

Reference materials. Participants were provided with a card listing keyboard command functions, and another showing flight information available through the simulated head-up display on the computer screen. Also, they were given a map and flight plan of their flight routes for both the practice and experimental missions. In addition, on the experimental mission they were given a reference, called a Tactical Standard Operating Procedure (TACSOP) book, which contained call signs, rules of engagement, and lists which differentiated between friendly and enemy military forces.

Data acquisition. The video from the computer simulation was recorded through a device, called a "Tvator,™" which converted the computer signal into a television signal, which was then recorded on videotape. Similarly, the audio signal from the intercom was transferred to the audio track of the videotape.

Scoring sheet. Videotapes were rated utilizing a scoring sheet developed to include the observable behaviors deemed to indicate good situation awareness (i.e., SALIANT). These behaviors were indexed to correspond to the legs along the route of flight to assist the experimenters in scoring the teams' behaviors. The behaviors could be rated as either positive, indicating good SA, or negative, indicating poor SA. Although most of the behaviors were positive, some negative behaviors were deemed to be obvious indicators of poor SA for the scenario developed. For each behavior exhibited by the team, they received one (positive) point for a positive behavior, and a negative point for a negative behavior. The sum of the points indicated overall team SA

In addition, each behavior corresponded to a dimension of situation awareness. Therefore, the measure was designed so that the total score could be broken-down into sub-scores that indicated how well the teams scored in each SA dimension.

Procedure

The participants were introduced to the experimenter and told they would be asked to fly a helicopter simulator as either the pilot or copilot/gunner on a military-style mission. They were then assigned to either position based on their overall flight hours, with the participant with the most hours being assigned as the pilot, and the other participant assigned to be the copilot/gunner.

Next, they viewed the videotape that explained how to operate the simulator, after which time they were allowed to ask questions. Following the videotape, the participants were given a practice flight in the simulator where they "flew" to a target range and were able to practice aircraft control and firing the weapons. During this flight the experimenter acted as the tower controller, and also answered any questions they had about the operation of the simulator.

Following the practice mission, the participants were shown the second part of the videotape, which presented a briefing of the scenario. In the scenario, the participants were to assume the roles of an Army attack helicopter crew. They were told they were one of five crews to participate in a reconnaissance mission to locate enemy armored vehicles and report if they found any on their route. They were cautioned that there could also be friendly vehicles in the same area. They were also

Table 1. Correlation Matrix.

	BAD BEH	COM 3	ORIENT 1	PROB 1	PROB 2	PROB 3	PROB 4	TEAM SAGAT
TARG PROSEC	-.624*	.538*	.085	.216	.611*	.534*	.152	-.092
WKLD 2	-.461	.574*	.272	.450	.189	.464	.467	-.039
WKLD 3	.377	-.373	.352	-.450	-.622*	.276	-.046	-.321
BAD BEH	1.000	-.297	-.177	-.620*	-.251	-.440	-.082	.065
COM 3		1.000	-.306	.096	.402	.266	.412	.064
ORIENT 1			1.000	.445	-.288	.386	.041	-.596*
PROB 1				1.000	.000	.209	-.190	-.314
PROB 2					1.000	-.100	.296	.353
PROB 3						1.000	.296	-.169
PROB 4							1.000	.481
TEAM SAG								1.000

*Indicates the correlation is statistically significant.

briefed on rules for attacking enemy targets (called Rules of Engagement), that stated they couldn't attack enemy vehicles unless they got permission, or unless the enemy fired on them first. The videotape included information such as call signs and controlling agencies.

Once they had viewed the briefing, they were instructed to don the interphone headsets and were given an opportunity to ask questions or conduct a team briefing. Since they were on interphone, their briefing was recorded on tape. They were further instructed that when they were ready, they could take the simulator out of the "pause" mode and start the mission.

During the scenario the experimenter acted the part of an external controlling agency. As the participants flew the route they were presented with several typical military aviation-related problems, such as discriminating between friendly and enemy forces, solving a loss-of-communications problem, and departing their planned route to search for a missing helicopter.

Twice during the mission, the experimenter stopped the simulation and administered a questionnaire to measure the participant's individual situation awareness, following the Situation Awareness Global Assessment Technique (SAGAT) developed by Endsley (Ref. 10). This questionnaire measured each participant's awareness of the current situation. The first one was administered after approximately the first one-third of the route was flown, and the second after about two-thirds of the route was completed. At the completion of this mission, the pilots were debriefed and paid for their participation.

3 RESULTS

Reliability

Cohen's kappa, correcting for chance agreement, was calculated for the ratings of two independent raters. Rater agreement ranged from .70 to .96 for the behaviors assessed by the SALIANT device.

Analyses

In order to test the extent that the behaviors assessed by SALIANT would be related to indices of performance and the SAGAT measures, a correlation matrix was calculated utilizing the following variables. Included in the matrix were 13 variables total. These variables represent the SALIANT dimensions (two orientation dimensions, three communication management dimensions, three workload management dimensions, and four problem resolution dimensions), total SA (i.e., the sum of all the SALIANT behaviors), two performance indices (i.e., targets prosecuted, and bad behaviors) teamsagat1, teamsagat2, and teamsagat (i.e., their total self-reported situation awareness using the SAGAT methodology. The SAGAT measures consisted of the average deviations of perception, per team, from the actual situation, that is, we calculated the average of the individual team members' perceptions. This correlation matrix is shown in Table 1 (only variables which had a statistically significant correlation are shown; the significant correlation is marked with an asterisk). The table shows that the correlational analysis indicated a significant positive correlation between problem resolution dimensions two and three with targets prosecuted. That is, the more frequent the problem resolution behavior, of the team, the more targets they were able to prosecute successfully. There was also a significant negative correlation between problem resolution dimension one and bad behaviors such that the more frequent their problem resolution behaviors, the fewer bad performance behaviors that were exhibited. There was also a significant positive correlation between targets prosecuted and communication dimension three, such that more of these communication behaviors was associated with higher numbers of targets prosecuted. Targets prosecuted and bad behaviors were also negatively correlated as well in the expected direction. Finally, a significant negative correlation was also observed between orientation dimension one and teamsagat, indicating that the more frequently teams engaged in these orientation behaviors the smaller their reported deviations from "reality" utilizing the SAGAT methodology. There was no significant, meaningful correlation of the workload dimensions with any of the other measures.

In order to test the extent that the SALIANT dimensions were significant predictors of the two performance indices relative to the total SAGAT score, two stepwise multiple regressions were conducted with bad behaviors and targets prosecuted as dependent variables. Results indicated that problem resolution dimensions one and three together predicted a significant proportion of the variance for bad behaviors. In addition, problem resolution dimensions two and three together predicted a significant proportion of the variance for targets prosecuted. These analyses predicted 58 and 73 percent of the variance respectively. Tables 2 and 3 depict the models for these two analyses. None of the other SALIANT dimensions nor the SAGAT variable were significant predictors of the two indices of performance.

Table 2. First Stepwise Multiple Regression Model

Model Summary

Model	R	R Square	Adjusted R Square	Std Error of the Estimate
1	.589 (a)	.347	.293	1.4539
2	.764 (b)	.583	.507	1.2138

(a) Predictors: PROB 1 (b) Predictors: PROB 1, PROB 3

Coefficients (a)

Model	B	Std Error	Beta	t	Sig.
1 (Const.)	4.585	.990		4.633	.001
PROB1	-2.146	.850	-.589	-2.526	.027
2 (Const.)	7.544	1.446		5.217	.000
PROB 1	-2.471	.721	-.678	-3.426	.006
PROB 3	-1.662	.666	-.494	-2.493	.030

(a) Dependent Variable: BADBEH

Table 3. Second Stepwise Multiple Regression Model

Model Summary

Model	R	R Square	Adjusted R Square	Std Error of the Estimate
1	.643 (a)	.413	.364	4.4494
2	.853 (b)	.728	.678	3.1652

(a) Predictors: PROB 2 (b) Predictors: PROB 2, PROB 3

Coefficients (a)

Model	B	Std Error	Beta	t	Sig.
1 (Const.)	7.304	2.209	-	3.307	.006
PROB2	5.826	2.004	.643	2.907	.013
2 (Const.)	-2.821	3.245	-	-.869	.403
PROB 2	6.359	1.434	.702	4.436	.001
PROB 3	6.128	1.719	.564	3.565	.004

(b) Dependent Variable: TARGETS PROSECUTED

4 DISCUSSION

The goal of the current effort was twofold. First, our goal was to develop a measure of team SA and second, to establish preliminary validation of the measure with one type of team. We sought to establish this preliminary validation by determining the extent that SALIANT would be significantly related to indices of performance, and also to test the extent to which the measure is related to an existing method of assessing SA (i.e., SAGAT), while still contributing unique variance to the prediction of performance. Results indicated some success in meeting these goals.

One of the first benefits of this approach is the ease with which raters can utilize the methodology and attain adequate reliability. Because the SA behaviors of interest are "scripted" to particular portions of the simulation or role play situation, it is relatively easy for raters to agree regarding the occurrence vs. non-occurrence of the behaviors. Obviously, however, the calculation of reliability must account for chance agreement by raters given the categorical nature of the data (e.g., use of Cohen's kappa).

Probably the most promising finding with regard to our results were the correlations of the problem resolution behaviors with both of the performance indices of interest in the current study. Specifically, the results indicated that higher rates of behavior associated with attention to the identification, prioritization, and resolution of problems were significantly correlated in the expected direction with targets prosecuted and bad performance behaviors (e.g., crashing, shooting friendlies, failing to interact as instructed with outside personnel). Furthermore, problem identification and resolution also proved to account for a significant portion of the variance in predicting bad behaviors, while problem prioritization and resolution accounted for substantial variance in the prediction of targets prosecuted. In contrast, the SAGAT methodology used failed to account for a significant portion of the variance for these measures. However, the SAGAT measures, obtained in the current study, were found to be related to the orientation dimension reflecting one's awareness of place in the mission. It might be that the SAGAT methodology is most useful for assessing the extent that teams are aware (or unaware) of their spatial position and place within their mission. Nevertheless, it is positive that although the two measures were related on this dimension, SALIANT still accounted for additional variance as well. As one would hope to expect, teams that managed to prosecute more targets were less likely to engage in "bad" performance behaviors.

Although the findings of this study are indeed promising, clearly this effort has established only preliminary validation for this methodology. The results regarding the problem resolution dimensions are particularly interesting. However, an additional point of interest of this study lay with the results regarding communication. To date, the majority of team research that has investigated team "process" behaviors has often focused on communication. In the current study communication regarding the verification of questionable information was related to the extent that teams were able to successfully prosecute targets. It appears that in answer to the question, can SA be identified as a construct separate from communication, the answer at this time is at least a qualified yes.

There are a number of issues that future research might address with regard to the utility of this methodology for the assessment of team SA. First, it would clearly be of interest to replicate the results of this study utilizing teams of a different nature. Specifically, it would be of interest to determine whether behaviors related to problem resolution are as useful in predicting performance of other types of teams. Second, this methodology was developed primarily to serve as a model from which situation appropriate measures of team SA might be developed for other types of operational teams (e.g., fire-fighters, medical teams). The success of the measure depends largely on its ability to serve in that capacity. In addition, although the problem resolution dimension of the measure did prove useful, as the literature review indicated previously, the SA construct has been considered in the past, to be one that is multi-dimensional in nature. Although there was one significant finding regarding communication, future efforts should determine the extent that SA is truly multi-dimensional and work toward improving the current methodology to better reflect those dimensions. However, we can consider possible reasons within the context of this study that we failed to do so as well as might be expected. For example, it might be that were no particularly meaningful findings regarding the relationship of orientation to the performance indices in large part due to the restricted variability with regard to this measure. That is, because trained pilots were utilized as subjects in this study, there was little variance with regard to behaviors related to disorientation in terms of their position in space. Although the dimensions adopted in this study were based on prior research, it would be most useful if the methodology could be tested utilizing a large number of teams in order to derive appropriate dimensions via factor analytic techniques.

Finally, a word of caution is in order with regard to the generation of SA behaviors to be scripted. It is critical that the behaviors reflected by the device be selected in order to minimize shared variance with performance, when performance is the criterion against which the behaviors are to be tested. This will obviously serve to prevent the inflation of measures of relationship calculated between the behaviors and performance. However, it must be noted that this is a highly important point. For example, few would argue that an aircrew that crashes in the absence of mechanical failure is almost certainly revealing a loss of SA. However, if crashes are used as an index of performance, this necessitates that crashes be disregarded as an index of SA.

In summary, the current study has found evidence that this methodology has some promise for increasing our ability to understand and measure the situation awareness construct as it applies to teams. Given the widespread use and importance of teams in today's world, it behooves researchers to identify and utilize methods that will increase our understanding of factors that improve team effectiveness. The use of teams is already so prevalent, the tasks they perform so varied, and the consequences for failure so often enormous that drastic measures are required in order to advance the state of knowledge in this area.

Acknowledgements

The authors wish to thank the pilots and instructors of Comair Aviation Academy, Sanford, Florida, and in particular Wayne Cenowa, Vice President of Operations, without whose assistance this study would not have been possible.

The views expressed herein are those of the authors and do not represent the official positions of the agencies with which they are affiliated.

References

1. Hartel, C. E., Smith, K., & Prince, C. (1991). *Defining aircrew coordination: Searching for mishaps for meaning*. Paper presented at the Sixth International Symposium on Aviation Psychology, Columbus, OH.
2. Salas, E. & Cannon-Bowers, J. A. (1997). Methods, tools, and strategies for team training. In M. A. Quiñones & A. Dote (Ed.), *Training for 21st century technology: Applications for psychological research*. Washington, D.C.: APA Press.
3. Endsley, M. R (1995). Toward a theory of situation awareness in dynamic systems. Special Issue: Situation awareness. *Human Factors 37*(1) 32-64.
4. Salas, E., Prince, C., Baker, D. P. & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. Special Issue: Situation awareness. *Human Factors 37*(1) 123-136.
5. Fowlkes, J. E., Lane, N. E., Salas, E., Franz, T., & Oser, R. (1994). Improving the measurement of team performance: The TARGETs methodology. *Military Psychology, 6*, 47-61.
6. Bunecke, J. L., Povenmire, H .K., Rockway, M. R. & Patton, M. W. (1990). The awareness component of cockpit resource management. *Proceedings of Psychology in the Department of Defense*, (pp. 259-263). USAF Academy Department of Behavioral Sciences and Leadership.
7. Leedom, (1990). Leedom, D. K. (1990). *Aircrew coordination training- A new approach*. Fort Rucker, Alabama: U.S. Army Research Institute Aviation R&D activity.
8. Prince, C. & Salas, E. (1993). Training and research for teamwork in the military aircrew. In E. L. Weiner, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management*, (pp. 337-366). Orlando, FL: Academic Press.
9. Schwartz, D. (1990). *Training for situational awareness*. Houston, TX: Flight Safety International.
10. Endsley, M. R. (1988). Situation Awareness Global Assessment Technique (SAGAT). *IEEE*.

PHYSIOLOGICAL PARAMETERS AS A POSSIBLE INFORMATION ABOUT COLLABORATION BETWEEN TWO CREW OF THE COMMERCIAL AICRAFT DURING LONG HAUL FLIGHTS

Lt. Col. Oldřich TRUSKA, MD, D.Av.Med.
Jiři ŠULC, MD, PhD
INSTITUTE OF AVIATION MEDICINE
P.O. Box 19, 160 60 Prague 6
CZECH REPUBLIC

SUMMARY:

By comparison of psycho-physiological response of a crew, consisting of two unequally experienced pilots it followed that the individuals with lesser amount of total and type specific flying hours than their counterparts are exposed to higher workload, regardless of their actual position within the crew. The cardiovascular response to the medium- and long-haul flights was more intense in flight officers than in commanders. The same difference appeared in subjective feelings of fatigue.

1. BACKGROUND

As a consequence of the scientific progress in aviation industry the multiple crews of wide body long-haul aircraft were gradually reduced to 2 – 3 persons. While modern avionics relieved the crew of substantial part of routine operations, biological variables, ensuing from the unequal degree of flying experience, differences in workload and the effects of environmental factors impose on the individuals in an immutable fashion. The share of roles between the two crew members of an aircraft follows from the operational routine, but at the same time the problem of an equitable distribution of workload as a basis of an effective collaborative crew performance remains somewhat contradictory.

A psycho-physiological strategy allows to study the impact of different type of stress on human subjects in operational context by means of non-invasive procedures. According to Roscoe (1) the use of psychological variables to assess workload is based largely on the assumption that they reflect the level of neurological arousal determined by the demands of the flight task, i.e. by workload. However, the credibility of data obtained should be augmented by other objective measurements and subjective ratings (2).

The aim of the study was to confront physiological and psychological responses of two-men crews of commercial wide-body aircraft in a series of medium and long-haul flights.

2. METHODS

The subject were 10 pilots, forming 6 crews of an Airbus 310-300 aircraft . Two pilots flew the mission twice. On partial legs of the trip the subjects changed proportionally in piloting from the left seat. Basic flight status, mean age and competence of pilots based on flying experience are summarised in the tablet No 1.

Crews were followed up during two sequential flights: the longer one with Prague and Singapore as the endpoints and a medium-range flight with Prague and Bangkok as the most remote places on route. The first and third layover stop in both flights were in the United Arabic Emirates and the second one at the place of final destination, respectively. From 10 to 14 days has passed between the departure and the return to the home base. The main work/rest characteristics of the routes are presented in the tablet No 2.

The Holter ECG monitoring served as the basic method to indicate the degree of cognitive demand through the mediation of heart rate measurements. Heart rate values, absolute and averaged for the two minutes ´ and one hour ´s readings were confronted with the actual activities of subjects (see the tablet No 3). The chronological records of all activities, performed by the crew, were registered by two cockpit medical observers, sitting in the jump-seats.

Two testing procedures were used for the fatigue evaluation . The first one was based on subjective fatigue rating in Bisson ´s inquiry (3),(see the tablet No 4) and the second one consisted in

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

the phoria determinations utilising the hand-held Maddox cross. Both methods were found useful in assessing human subjects in demanding aviation environment.

The data regarding both groups (i.e. the commanders and co-pilots) was analysed using the T-test and one-way ANOVA. The level of significance was set at $p < 0,05$.

3. OBSERVATION

Statistically significant differences between captains and co-pilots in the total flying hours and Airbus hours were proved.

Reliable psycho–physiological data from 46 partial legs was obtained.

Pilots spent in practical terms nearly all flight time in the cockpit. They left their seats one to maximum three times for only a very short time (less than 15 minutes) to relieve nature or to perform a short leg exercising. Even in restful parts of the horizontal flight the activity of co-pilot was conspicuously more intense than the activity of captains, while their unrest did not depend on actual functional position in the crew. They repeatedly skimmed the on board documentation and pilot ´s manual and controlled the navigation computer. They also performed more numerous moves to find optimal position in the seat and stretched their trunks and limbs more energeticly than pilots on the adjacent seat. The difference in global motor activity follows from the examination of on-board medical observers and could not be supported with the actigraphy.

Instantaneous values measured during the entire flight displayed high stability, except for the take-off and landing phase. Despite the fact, that individual minimum and maximum heart rate values in eastward flights were higher than in westward flights, the difference was not statistically significant. The minimum heart rate value kept down during monotonous parts of the flight even for tens of minutes, whereas the highest heart rate values has kept for only a few seconds.

Absolute values as well as the relative increment of heart rate were higher in co-pilots compared with captains. In both eastward routings the relative heart rate increment in co-pilot ´s group was by 13,1 to 15,1 % higher compared with captain ´s group (Figure No 1A, 1B, 2A,2B). During westward flight the difference decreased to 3,7 – 8,5%. Considering the operative position of the subject in the crew, i.e. the differentiation between the handling and non-handling pilot, the cardiac response did not change substantially. The relative increase of heart rate in handling or non-handling captains was always (with one exception) milder than that of handling or non-handling co-pilots, while the degree of activation response was stronger during eastward than westward flight.

Effect of cumulative fatigue manifested itself in the progressively growing scores in the Bisson´s test (see figure No 3). Compared with the phoria measurement the subjective fatigue ratings were more consistent and productive. The scores seen higher in co-pilots, did not show significant differences between the crew members. Signs of sleepiness in both pilots were also registered during the boring phases of particular flights. Their dynamism was analysed in detail in a paper, presented recently elsewhere, thus no particular attention will be paid to this physiological phenomenon.

4. EVALUATION

Some representative papers, dealing with the in-flight activities of the members of multiseat crews arrived at different conclusions as to the projection of mental efforts, workload and stress into the measurable psycho–physiological response. The prevalent opinion is, that pilots in control of two–pilot aircraft have higher heart rates, than co-pilots (2). When using heart rate to support subjective ratings, more reliable results can be expected for handling pilots than for flight officers (1). Hart and Hauser (5) have found significantly higher heart rate for aircraft commanders of the C-141 "Stargazer" than for the co-pilots. The differences were particularly striking when it is remembered that the pilots were fully qualified in both position. The correlation among heart rate, effort ratings, stress ratings and workload ratings for all pilots regardless of position, were significant, indicating, that similar factors affected both the subjective and physiological responses. When computed for each position individually, however, the correlations mentioned were considerably higher for the aircraft commanders than for the co-pilots. This suggests that the stress associated with the responsibility for piloting the aircraft affected the command pilot´s unconscious responses to stress, while this was not the case for the co-pilots.

Kakimoto´s and Nakamura´s (6, 7) teams in a series of measurements of changes in heart rate, cortisol from saliva and catecholamine excretion in crew members of C-1 transport flights have also confirmed, that the activation level was significantly affected by the degree of responsibility for piloting.

In addition to it they revealed the impact of flying experience upon the vegetative and hormonal response: the commanders displayed always milder and more ballanced reactions.

The results of present study differ in one aspect from generally accepted findings. The measures of activation in flight officers, irrespective of their function in the crew, signalled more effort exerted by them to fulfil the professional demands. Not only the heart rate increment was systematically higher than for the commanders, but also the general behavioural activity reflected an ambition to pass muster. Captains occasionally penalised this behaviour, even when sitting on the right seat. The experience of an increase in workload does not imply an improvement in performance, but it can contribute to the cummulative effect of fatigue, more intensively perceived by flight officers. The source of an immoderate psycho – physiological activation follows from the unbalanced flying experience between the subjects and should be considered in Cockpit Resource Management training.

3. References

1. Roscoe, A.H.: The practical assessment of pilot workload. (1987). AGARD-AG-282.
2. Caldwell, J.A.(Ed.): Psycho-physiological assessment methods.(1994). AGARD-AR-224
3. Bisson, R.U., Lyons, T.J., Hatsel, C.: Aicraft fatigue during Desert Shield C-5 transport operations. (1993). Aviation. Space Environ. Med., 64, 848-853
4. Šulc, J.,Vanko, J., Bartoš, D.: Fatigue in Airbus crew members on long haul flights. (1995). 43rd ICASM, London.
5. Hart, S.G., Hauser, J.R.: Inflight application of three pilot workload measurement techniques. (1987) . Aviation Space Environ. Med.,58, 402-410.
6. Kakimoto, Y. et al.: Crew workload in JASDF C-1 transport. flights. I. Change in heart rate and salivary cortisol. (1988). Aviation Space Environ. Med., 59, 511-516.
7. Nakamura, A. et al.: Crew workload in JASDF C-1 transport flight. II. Change in urinary catecholamine excretion. (1989). Aviation Space Environ. Med., 60, 301-306.

Tablet No 1, Sample description

		Flying experience (hours)	
Flying status	Mean age	Total	Airbus
Captains	48,3	11 271	589
Flight engineers	45,1	4 434	235
p	n.s.	0,05	0,01

Tablet No 2, Main work/rest characteristics

Routings	*Pilots*	*Mean Flight time (h: min)*	*Layover (h)*
PRG – SHJ		5 : 47	73
SHJ - BKK*- SIN	6	7 : 15	76,5
SIN - BKK*-AUH		7 : 49	25
SHJ – PRG		6 : 17	-
PRG – AUH		5 : 05	91
AUH – BKK	6**	5 : 26	84,5
BKK – SHJ		6 : 10	97
AUH – PRG		6 : 13	-

* Technical landing in BKK only (70 – 110 min.)
** One cockpit crew not monitored on AUH – BKK and AUH – PRG routing

Tablet No 3, Methods for workload assessment

Method	Stage of flight	Reliability
Monitoring of behaviour	Continually	+ + +
Holter ECG monitoring	Continually	+ + +
Fatigue assessment		
BISSON' s inquiry	Before starting and after stopping the engine	
MADDOX CROSS test	1 hour after take-off	+ + +
	2nd	
	1 hour before landing	+

Tablet No 4 , Subjective fatigue rating scale

BISSON et al. 1993	
1.	FULLY ALLERT
2.	VERY LIVELY , RESPONSIVE
3.	O. K. SOMEWHAT FRESH
4.	A LITTLE TIRED
5.	MODERATELY TIRED, LET DOWN
6.	EXTREMELY TIRED, VERY DIFFICULT
7.	EXHAUSTED, UNABLE TO WORK

Fig. 1A: Relative HR Increment
(mean 1 hour's values)

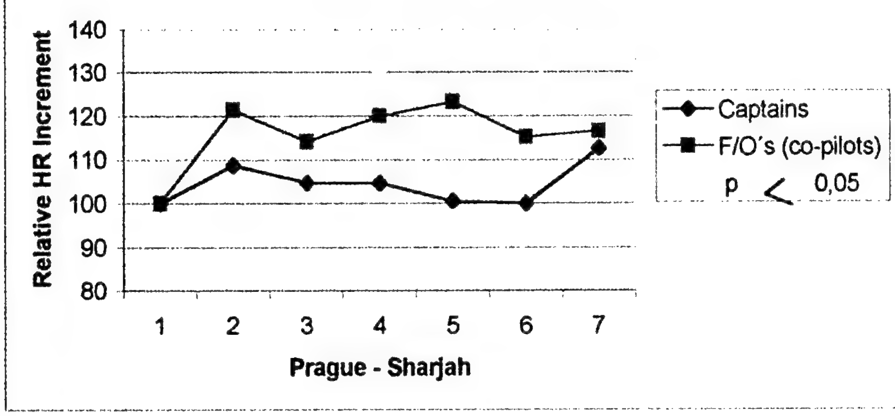


Fig 1B: Relative HR Increment
(mean 1 hour's values)

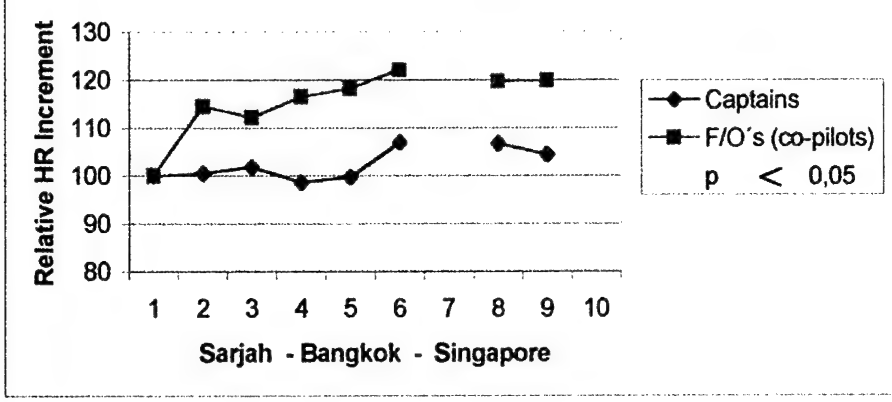
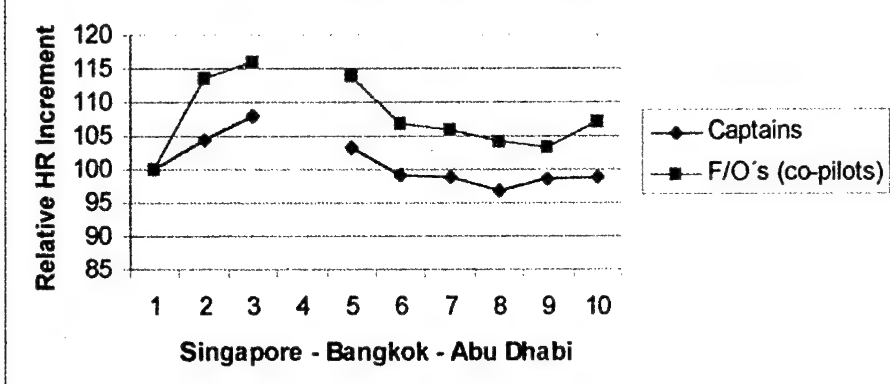


Fig 2A: Relative HR Increment
(mean 1 hour's values)



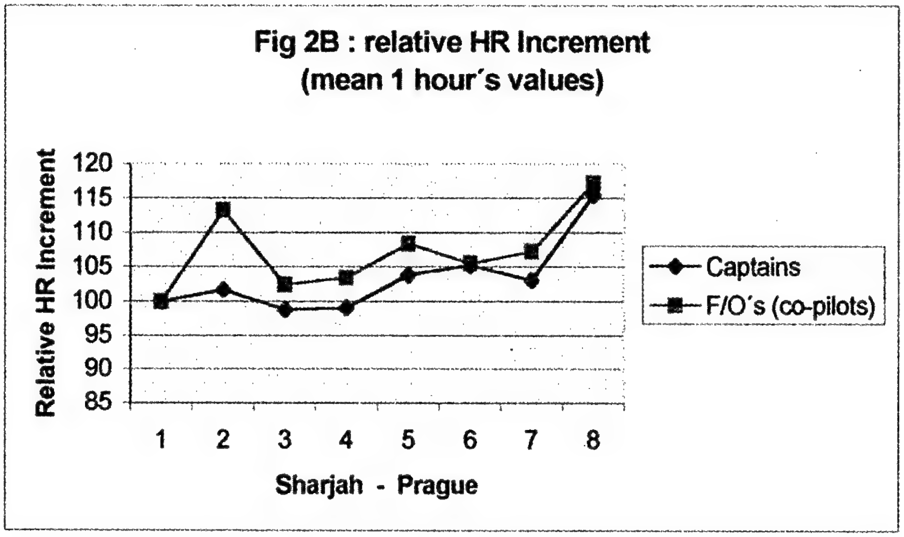
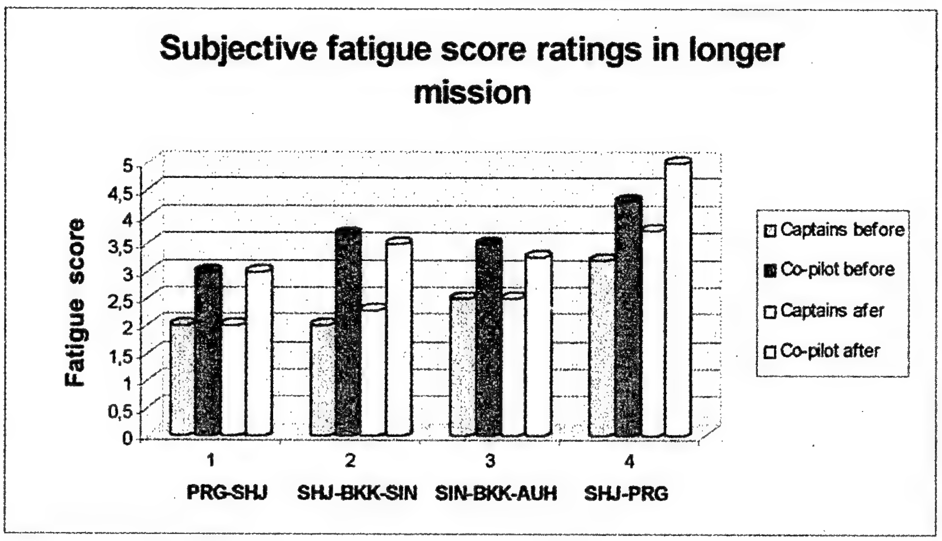


Fig 2B : relative HR Increment
(mean 1 hour's values)
Relative HR Increment
120
115
110
105
100
95
90
85
1
2
3
4
5
6
7
8
Sharjah - Prague
Captains
F/O's (co-pilots)


Subjective fatigue score ratings in longer mission
Fatigue score
5
4,5
4
3,5
3
2,5
2
1,5
1
0,5
0
1
PRG-SHJ
2
SHJ-BKK-SIN
3
SIN-BKK-AUH
4
SHJ-PRG
Captains before
Co-pilot before
Captains afer
Co-pilot after

CSCW: Towards a Social Ergonomics?

Liam J. Bannon
Interaction Design Centre
University of Limerick
Room FG005
Limerick
Ireland

1. SUMMARY
This paper argues that there is a need for a richer understanding of the sociality of work in the Human Factors field. Some have characterized this shift in perspective as a move from a cognitive to a more social ergonomics. Problems with the cognitivist approach in the human-computer interaction (HCI), and, more generally, the human factors, research programmes, are outlined. The emerging field of Computer Supported Cooperative Work (CSCW) is discussed. The ethnographic studies of workplace practice which have been performed in CSCW are shown to provide another perspective on communication, collaboration and cooperation in the workplace which should be of interest to HF practitioners and complex systems designers. Efforts to open up dialogue between the HF and CSCW communities are noted, and possible new hybrid frameworks discussed.

2. INTRODUCTION
Ladies & Gentlemen, thank you for the invitation to make a presentation at this conference. I must admit, that when I was first contacted by Michael McNeese concerning this meeting, I was somewhat surprised, as my work has not directly involved studies of C^3I, (although I have recently begun some studies of air-traffic controller (ATC) coordination), and so at first I was unsure as to exactly what I could contribute to your deliberations. However, after further discussion with Michael, we agreed that perhaps having someone address you from "outside" the mainstream human factors (HF) community could be beneficial, especially to open up some discussion concerning on-going developments in the Computer Supported Cooperative Work (CSCW) field, and their potential relevance for this community.

The title of the talk refers to my own perspective on the evolution of the human factors field, moving from a focus on physical to cognitive, and perhaps now, social, aspects of the human -technology "fit". The intent is not to present the CSCW field as having all the answers to the latter aspect of ergonomics, but to initiate some dialogue between HF practitioners and CSCW researchers on items of mutual interest. (This notion of a "social ergonomics" has been influenced by discussions with John Seely Brown of Xerox PARC, back in the early 1980's. See his chapter: *From Cognitive to Social Ergonomics and Beyond* in D. Norman & S. Draper (1986) (eds.) User Centered System Design: New perspectives on human-computer interaction).

The main thesis of this talk can be stated thus: *That much current human factors work embodies an unnecessarily restrictive perspective on the nature of the human - technology relationship, and that encounters with alternative conceptualizations than the prototypical "information processing" model of the human could provide new insights into the ergonomics field.* The specific issue which I will discuss here is the way in which an understanding of what might be termed the "sociality" of work can provide new vistas on the design of mediating technologies for complex work settings. What I will argue for is the need to continually critique and re-invent our objects of study. The issue is not simply the need for an accretion of disciplinary perspectives e.g., cognitive, social, organizational, but rather an openness to re-visiting our foundational assumptions concerning the nature of human interaction - with others, and with (or through) artefacts.

The structure of the paper is as follows: I will start out by saying a few words about my own academic background. Then, I briefly review some arguments about how the human factors area, and more specifically the human-computer interaction (HCI) area, has evolved, with both recently taking a cognitivist perspective. This perspective has itself come under scrutiny in recent years, and one can discern attempts to introduce new conceptualizations and methodologies to the field. I will mention some of these approaches, and then focus on work within the CSCW field, which has been heavily influenced by

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

ethnography. After noting some of the insights that this approach has produced, I will attempt a rapprochement between certain kinds of ergonomic work and the CSCW field, and note some efforts in this regard. Whether this pluri-disciplinarity affords new insights or simply results in an "unscientific" and vacuous *melange* is another matter, of course, and fierce arguments rage in certain quarters about this, although my own position leans to the former. Finally I will note some recent fora where these different perspectives and communities have met, hopefully with benefits to all of the differing perspectives.

3. Some Personal History

I was trained in psychology and computer science in the early 70's , and became interested in the use of computational models in psychology, as well as in human aspects of using computers. As a doctoral student in cognitive psychology in Canada, I continued these interests, and also spent a year in Minneapolis with the Honeywell Systems and Research Centre as one of their first 3 human factors interns in 78-79. I worked with Don Norman on the UCSD HMI project 82-85, and was influenced by many people and disciplines there, including Mike Cole in the LCHC Lab, Bud Meehan and Aaron Cicourel in sociology, Roy d'Andrade in cultural anthropology, and many others. My time at UCSD had a formative influence on my subsequent work. Subsequently I spent time in Scandinavia with people involved in participatory design (PD), especially the Aarhus school - notably Susanne Bødker, Morten Kyng, and Pelle Ehn, also attempting to apply cultural-historical activity theory in order to better understand the relations between people and technology. It was during the late 80's that I became involved in the CSCW field, and I have had a major involvement with the European CSCW Conferences from their inception in 1989, and with the CSCW Journal, launched in 1992. Currently in the University of Limerick Interaction Design Centre, I am involved in a variety of HCI, Multimedia and CSCW projects. While I have not been heavily involved in the aviation area to date, I do have contact with Capt. Neil Johnston, of Aer Lingus, that many of you may know, an associate editor of the Journal of Aviation Psychology, with extensive experience in pilot training and particularly in Crew Resource Management (CRM), and I have had close links with a number of UK sociologists concerning their ATC studies, organizing a symposium on differing approaches to ATC - behavioural, cognitive, sociological, at the WEAAP Conference in 1994 in Dublin (1). I am also currently involved in studies at the ATC Centre in Shannon airport in Ireland, focusing on issues of mutual awareness and differences in work activities across ATC centres in Europe. So, while I was initially trained as a cognitive psychologist and computer scientist, my more recent work owes much to sociological and socio-cultural perspectives in understanding human activities and the mediating role of technologies. Let me now turn to provide a brief account of some concerns I have had with the HCI work in the late 80's and into the 90's, which I believe also affect much general human factors work in this period.

4. Crisis in HCI?

It has often been observed that the HF field has moved from a concern with physical ergonomics to a more cognitive ergonomics. The old field of "man-machine communication"(sic) has given way to HCI. Of course many other interests are involved in the general HCI field, but I would argue that for many HF personnel, HCI would be seen as a central preoccupation. Within HCI, the cognitivist perspective has been dominant for well over a decade, although it has been under attack more recently. Since I have written extensively elsewhere concerning the critique of cognitivist HCI (2,3,4), I will not go into these arguments in any depth here, but simply note a few of the concerns raised.

The basic thesis is that this mainstream "cognitive science"-inspired HCI research has come up against a number of problems, both in terms of its conceptual frame, its research agenda and the usability and utility of its empirical results for the software design community. Emphasis has been on the individual user's model of the task, the actual behaviour of users, their errors, etc. Much HCI modelling work is still undertaken with a view to replacing human skill by "intelligent" systems, rather than with the intent of supporting people via the design of better computer tools. Advances in HCI seem to emerge from design groups without any clear lineage from the conceptual frameworks or empirical methods touted by cognitive science. Early research done in the HCI field was confined to rather small controlled experiments, with the presumption that the findings could be generalised to other settings. It has become increasingly apparent that such studies suffer from a variety of problems that limit their usefulness in any practical setting. The social nature of much human learning is downplayed. Questions of motivation in the performance of experimenter-defined tasks are not considered sufficiently. The underlying model of the "user" apparent in this perspective seems at times patronising and misguided - naive users, idiot-proof system design, etc. That there is some form of crisis within segments of the HCI community can be gleaned by the emerging debate about the role of cognitive theory in HCI (5), the increased emphasis on usability issues, and on the expressed need for field studies. There is a call for changes, of a more or less radical kind, in the conceptual frameworks employed, the kinds of research undertaken and methods used in HCI. For example, Thomas & Kellogg (6) discuss the

"ecological gaps" caused by bringing studies into the lab, both by omission of factors in the real world, and by the addition of new elements in the testing situation that do not correspond to real world eventualities. Perhaps one of the most important kinds of gaps noted, in the present context, are what Thomas & Kellogg refer to as "work-context gaps" concerning the social setting, the culture of the workplace, etc. It is this lack of appreciation of the use setting that is a major problem with much of the cognitive science-inspired HCI work to date.

There have been a number of attempts to answer some of these criticisms, both from within the cognitive tradition and outside it. From within, the work of Hutchins (7) on "distributed cognition" is a bold attempt to keep many of the concepts from cognitive psychology - such as computation and representational systems - but apply them in novel ways to situations, showing how several human actors and artefacts can be viewed as "propagating" representations. This work is also distinguished by insightful ethnographies of work practice. The information processing view of how people function is also being enriched, if not replaced, by a range of perspectives such as symbolic interactionism, ethnomethodology, cultural-historical theory, and phenomenology, which require us to re-examine the ontologies and epistemologies espoused in traditional material. For example, Winograd & Flores (8) provide a radical critique of the Western intellectual tradition, and explore an alternative perspective based on biology, hermeneutics and phenomenology. Lave (9) has provided a powerful critique of accepted views of learning and practice, noting that "cognition observed in everyday practice is distributed -stretched over, not divided among- mind, body, activity and culturally organized settings (which include other actors)". Others have found inspiration in the materialist philosophy expounded by Marx and Engels that emphasizes praxis as the basis for human development. The work of the Russian psychologists Vygotsky and Leontiev are examples (10, 11, 12). More recent work in this tradition has direct implications for HCI (13, 14).

Recent work in the area that has become known as "social studies of science and technology" has also had an impact on how we view the evolution and use of artefacts in human activity. This approach emphasizes the way in which technology itself can be viewed as an actant in networks of humans and non-humans. While this "actor-network theory" may appear strange on first sight, one cannot but be impressed with the insights about the relationships between people and technology therein (cf. e.g. Latour, 15). Yet another critique of the cognitivist approach comes from ethnomethodology, focusing on the situated nature of human action in workplaces and its consequences for models of the user and ultimately, software design. The publication of Lucy Suchman's book *Plans and Situated Actions* in 1987 can be seen as the landmark publication of this genre. It was Suchman's book which introduced many people involved in software development and human-computer interaction to ethnography and more specifically, arguments from the field of ethnomethodology, in a powerful critique of plan-based models of human behaviour, much in vogue in the field of artificial intelligence at the time, and influential in both cognitive psychology and the HCI field.

Concomitant with these concerns was one questioning the human-computer dyad as the fundamental unit of analysis. This questioning was coming from many quarters, including those investigating organizational change, economic modelling of the firm, and from field studies of workplaces. The emergence of the field of CSCW was one development around this time (mid -eighties), where there was a conscious shift in focus towards the need to understand how people accomplish their activities cooperatively in the workplace. Yet another approach which also began to receive more coverage at this time was the ongoing work, mainly from Scandinavia, concerning participatory design practices in systems development (16).

Methods as well as theories in HCI have come under critique. For example, Landauer (17,18,19) has decried the poverty of many of our experimental manipulations, and attempts to push psychology out of the laboratory setting in order to be more directly relevant to human needs in the workplace. He notes: "*There is no sense in which we can study cognition meaningfully divorced from the task contexts in which it finds itself in the world*". Yet, this admission is often not followed in practice. Carroll and his colleagues (20) have elaborated on the "task-artefact cycle" as an attempt to approach the issue of the co-evolution of artefacts and human tasks. Whiteside & Wixon (21) give some nice examples of how far removed some of the cognitive science work is from real world situations. It is this lack of appreciation of the use setting that is, in our view, a major problem with much of the cognitive science HCI work to date.

5. Winds of Change in Human Factors ?

Within the Human Factors field more generally, while the cognitivist paradigm has had a major impact, there has still remained more of a focus on actual work tasks and on larger issues of human-machine interaction, than that evident in the field of HCI. Despite the drive towards greater automation and the use of expert systems, one can also detect a movement against the prevailing orthodoxy concerning the human being as being the weakest link in complex

systems. This alternative view, which cautions against over-automation, seeks to re-habilitate the human as a central - frequently positive - component in complex human-machine systems. It can be found in a variety of HF literature over the past several years. Examples include: Bainbridge's early paper on "ironies of automation", Wiener's critique of the "glass cockpit", the work of Woods and Roth on the limitations of expert systems in use, and Hollnagel and Woods development of the cognitive systems engineering paradigm, where they view computers as instruments not prostheses. All these approaches reject the exaggerated claims of artificial intelligence, and return more power and control to human operators in complex human-machine systems. It is important to stress that such "human-centered design" (a term not without its difficulties) is being advocated, not simply from a particular value framework concerning human working conditions, but also as a result of many failures in reliability of advanced automation facilities. The flexibility and adaptability of people, their "sense-making" capabilities (Weick (22)) are difficult, if not impossible to achieve with so-called expert systems.

One area of increasing interest in HF is that concerning crew, team, or group performance, as evidenced, for example, by this particular conference. Much of the earlier work on HF has focused on the individual operator and his or her tasks. Issues of communication, collaboration and coordination, while central to concepts such as C^3I, have not benefited from a strong conceptual paradigm. As already noted, one of the striking features of the cognitivist paradigm has been its focus on what goes on "in the head" of the individual, neglecting the central role of external artefacts and other people in the accomplishment of work activities. Many see the birth and evolution of the computer-supported cooperative work (CSCW) field since the mid-eighties as an expression of this newly emerging set of concerns. The appropriate conceptual frameworks and methods for studying this new area are open to some debate. Some see the field as merely a simple extension of HCI, and thus argue that existing tried and tested methods and concepts from cognitive psychology are all that is required. Others argue for a more radical approach, claiming that the "cognitivist" approach needs to be abandoned as it is not only perceived as conceptually flawed, but unable to deal with the complexities of the workplace. They see sociologically-inspired concepts and methods as being more appropriate. In particular, CSCW has turned to ethnography (often ethnomethodologically-inspired) as a key approach towards understanding the sociality of work. Let us look briefly at this developing field, and see what it may have to offer, concerning these questions of cooperative work and its support.

6. CSCW - a turn to the social?

The term "Computer Supported Cooperative Work" (CSCW) has come to embrace a variety of research in such overlapping areas as workgroup computing, collaborative computing, groupware, co-ordination technology, augmented business teams, group decision support systems, and cooperative work support. One of the key features of this new field is an interest in supporting groups or ensembles, rather than individuals or whole organizations, with information technology. CSCW entails both a wider remit than traditional information systems as regards the different settings in which it is appropriate to study cooperative work arrangements, as well as a more explicit focus on the "support requirements" of cooperative work and the way people create, manage, disable, modify, etc. computer-based mechanisms of interaction than is seen in the other fields. Studies in such diverse areas as computer-aided design (CAD), computer-integrated manufacturing (CIM), computer-aided software engineering (CASE), group decision support systems (GDSS) etc., are all relevant to the CSCW field to the extent that they study the use of computers to support cooperative work in different domains. A focus on the multiplicity and complexity of cooperative work arrangements and problems and prospects for their computer augmentation is what some would regard as what is "new" in the field. The importance of articulation work (23) - the work that is required to be done in order for any division of labour to function smoothly - in understanding how people manage to co-operatively accomplish their work is another insight that has been investigated in some detail in CSCW work (24).

It is not my purpose here to provide an overview of the whole field of CSCW, rather, I simply wish to bring to your attention the accumulation of workplace studies that provide insight and understanding of cooperative work activities, and focus on the sociality of work and the consequences of taking this notion seriously in designing computer support. The lack of any such CSCW studies at this conference - devoted to collaborative crew performance in complex settings - is striking. In what follows, I will attempt to show the differences in perspective provided by these mainly ethnographic studies of the workplace, compared to traditional HF task analyses, and move on to discuss ways in which our different approaches may be mutually beneficial.

Fundamental to much CSCW is an ethnographic approach to field studies. This contrasts strongly with the normative task-analytic approach to understanding human activities in much HF work. Ethnography is not simply about going out into the field to collect data, more importantly, it is an analytic framework for the research. As Anderson (25) notes: "Within the social sciences, ethnography is a form of reportage

and not a form of data collection......the ethnographer's eye is always intepretive". The ethnographic work provides a focus not simply on actual work situations, but also describes them in ways which reformulate the relationship between researcher and object of study - "the analytic value of rendering strange and/or exotic some of the key taken-for-granted features of the setting." (26). There is debate as to the exact relevance of these studies to system design per se, as initially some people viewed such studies as producing "requirements", which is far from being the case. However, the current situation could best be characterised as one where many in the field accept the relevance of these ethnographies in exposing the artful ways in which people "get the work done" in spite of breakdowns and crises of various kinds, their ability to cope with constant interruption, the ways in which local knowledge is used to shape the work in a matter-of-fact and unremarked-on fashion, the importance of "mutual awareness" in many complex work settings, the need for people to "gear into the work", the interweaving of individual and collaborative work, etc.

More generally, Hughes and King, (27) note some of the framing concepts that underlie the ethnographic approach used in these workplace studies: "assume from the outset that the world is socially organised; see the world as socially organised from within the setting; understand the work and its activities in terms that member's understand and use; go into the work setting and examine work activities in all their detail; treat work activities as part of the flow of work; don't treat domains as equivalent; don't draw a distinction between expert knowledge and practical knowledge; don't draw a sharp distinction between activities and technology; don't classify users." All of these insights have become part of the background against which the CSCW community discuss computer support. They provide a perspective that has helped to illuminate the workaday world and should serve as a backdrop for anyone that wishes to develop computer systems. In order to make this more concrete, I provide short vignettes on how this approach might provide a different perspective to more standard HF views in the case of ATC work, and the allocation of functions topic.

6.1 Vignette: The Work of Air Traffic Control

Air Traffic Control (ATC) is a very complex work activity, involving a variety of tasks that must be interleaved, with obvious time constraints, including communication and cooperation, not just with cockpit crew, but with other controllers on the ground. Given the safety critical nature of the activity, there has been a strong interest in understanding and either automating or supporting the work of the controllers with computer aids for many years. Failures of development efforts that relied on extensive use of automation have influenced designers and decision-makers. A more considered, evolving piecemeal strategy for supporting the work of ATC officers with technology is now developing among a number of researchers in the area. There remain however serious questions about human-machine task allocation in the design of the supporting technology systems, for ATC, just as for other areas involving intense monitoring of workflows and processes. Reducing the involvement of controllers in the actual controlling activity through automation can have the effect of making them less competent as controllers when in fact a system malfunction requires them to intervene and assume direct control for a period .

The work of air traffic controllers has been the subject of study for many years by human factors researchers. As the work involves a high degree of expertise, it has been evident for some time that extrapolation of simplified laboratory studies to the world of work is deeply problematic in this area, so there has been a tradition of performing studies in as rich a setting as possible, and using experienced people in the studies. The conceptual frameworks employed in such studies have been varied. As well as a behaviourist task analysis approach, there has been an interest in what has been termed "cognitive task analysis" (CTA). This differs from traditional task analysis, which tends to focus on behavioural performance and training objectives, by putting more emphasis on the cognitive processes involved in producing behaviour and on the nature of the learning process. There has also been an increased interest recently in studies that focus more explicitly on the *socially organised character* of ATC work, and the effects that new technology might have in supporting or indeed hindering this collaboration. Much of this work has been reported within the area of CSCW - Computer Supported Cooperative Work, and is often identified with an *ethnographic* approach to studying ATC work, *in situ.* This is quite distinct from the individualistically-oriented task-analytic studies done in laboratories, as often found in traditional human factors work, and while as we shall see, there are overlaps between this approach and CTA, there are also quite large differences in approach. As expressed by Hughes et al., (28?): "There is no one method of ethnographic analysis. ... The field workers immersed themselves in the work by spending several months observing activities on and around the suites, talking to staff, and discussing with them the researchers' developing understanding of what controllers do. While attempting to avoid prejudices and to allow the work situation to 'speak for itself' as much as possible, researchers cannot claim to address it innocent of any theoretical orientation; and their results would be much impoverished if they did. The purpose of an ethnographic approach is not so much to show *that*

work is socially organised (which is rather easy) but to show *how* it is socially organised."

The ethnographic studies note, for example, how much individual work done by ATC officers is done in such a way as to make available to other ATC officers information concerning the state and conduct of the task, without requiring overt messaging between officers. In any attempt to "improve" the tools of the ATC officers, they point out how great care must be exercised so that the current, often non-obvious, fluid use of artifacts and signs to communicate the state of affairs in the airspace to all in the control room are not disturbed by new technology. The technology may, while making specific tasks easier, have the undesirable side-effect of occluding some vital information about the state of affairs to others. An example that has been studied in this regard is how properties of the physical flight control strips are used to make available information about the state of affairs to others in simple yet undemanding ways that would be difficult to replicate in simple electronic replications of such strips. This has lead to a re-examination of how to provide opportunities for mutual awareness through the new technologies. Our own recent work at Shannon ATC Centre is examining the ways in which the controllers maintain mutual awareness through the use of implicit modes of communication, such as peripheral monitoring and overhearing conversations (29). As developed in (30), these communications are mainly supported by the existence and extensive use of available resources provided in the workspace (shared radar screen, visible strip board and audio access to radio communication). One interesting extension of our work is the comparative analysis of French and Irish ATC activities (31) which is investigating how, despite major differences between the work settings (both technical and organisational), French and Irish controllers similarly elaborate and update their mutual awareness in order to ensure the efficiency and safety of their activity.

6.2 Vignette: Re-visiting the "Allocation of Functions" issue

In a recent keynote address at a conference that I helped organize on the theme of new perspectives on the allocation of functions (could we have a more sacred HF concept?) John Bowers (32) presented a CSCW perspective on this hoary concept,. The kinds of queries he raises provide some idea of how the CSCW field may be able to cast a somewhat different light on traditional HF topics. Based on ethnographic studies of workplace practices, Bowers asks - do people perform "functions" at all (in terms of input-output relations)? Does work come ready-sliced (so that it can be meaningfully allocated in pieces)? He distinguishes between "allocation from without vs. from within" concerning the way in which people coordinate their activity *in situ*, as distinct from having a division of labour imposed externally, arguing that we should be supporting a working division of labour rather than encoding a process model. He emphasizes what might be termed "the work to make it work", i.e. the articulation work required of the workers in order to mesh their tasks and ensure the smooth flow of the overall activity. This requires that we try to support mutual awareness, rather than attempting to decompose tasks, and emphasizes the multiple participation roles in human interaction, not just sender- receiver communication, which ensure that work groups can collaborate effectively through possibilities of overhearing broadcast information in work settings.

6.3 Summary

What I hope to have achieved in this Section is simply to note that research within the CSCW field, especially the corpus of ethnographic and ethnomethodological studies of work, provide a rich resource for understanding human activity, especially concerning aspects of communication, collaboration and coordination, which one does not find within the usual HF tradition of task analysis - behaviourist or cognitive. Concepts such as articulation work, peripheral awareness, working division of labour, situated action, coordination mechanisms, common information spaces, have been exploited in the CSCW field and provide a rich set of conceptual resources for discussing the nature of cooperative work, which I believe should at least be discussed within the HF community. Let me now turn to how this commingling of communities may be facilitated.

7. The way forward

I have noted above some of the CSCW contributions to the understanding of cooperative work that may be of interest to the HF field. At the same time, I would also like to note that there are bodies of work within the HF arena that are not well known to the majority of the CSCW community that would be in turn informative for CSCW. While much work in cognitive ergonomics and cognitive engineering is aligned with the cognitive-science inspired HCI tradition that has been criticised above, it is also important to recognize the pluralism that exists within these areas. Specifically, I would like to note two strands of work, one associated with Francophone cognitive ergonomics and studies of the course of action of human activities, the other with cognitive systems engineering, that involve detailed studies of human activities and work practices which have often not been given due notice within the North American-dominated HCI arena, nor perhaps, in CSCW.

The first body of work which I refer to is that of European, mainly Francophone, researchers such as Maurice de Montmollin, Jacques Thereau, Veronique de Keyser, etc., all of whom stress the importance of conducting field studies, and who are fully aware of the distinction between actual work practices and normative accounts of work. A useful starting point for understanding this approach is the summary paper by de Montmollin (33). There, de Montmollin argues : "Operators actual activities have to be distinguished from the tasks they are requested or supposed to perform; operators working in natural life environments have to be distinguished from anonymous and universal human beings; complex natural life environments have to be distinguished from the interfaces, as the whole has to be distinguished from one of its parts". Such sentiments would appear to fit in very well with the background orientation and observations conducted by ethnographers. Indeed, de Montmollin goes on to note: "ergonomics analysis and modelling of activities cannot be anything but natural field analysis, in an ecological perspective". Again, this approach has resulted in numerous field studies and reports which illuminate in striking ways the complexities of everyday work activities, and the role of artifacts in their accomplishment.

The second body of work also has a strong, but not exclusively European flavour, characterised again by a marked emphasis on the necessity for field studies, but also by a more systems-level analysis, under the rubric of "cognitive (systems) engineering". While again the work of such figures as Rasmussen, Hollnagel and Woods is well known in ergonomics fields, especially the IFAC community, it is not nearly so prominent within the broader HCI community, as evidenced by publications or citations of this work in, say the CHI or CSCW conferences. This work is also characterised by detailed field studies, by attention to the flow of activities in the workplace and the use of artifacts as tools and media, and importantly, by a concern for how these studies can inform the design process.

The argument here is that many researchers in CSCW have tended to dismiss the work of psychologists in the HCI field for ignoring crucial aspects of work organization and the work setting, and critiques the HF field for too narrow a perspective on individual tasks in human work. While such criticisms are indeed justified against many HCI and ergonomic studies, they do not necessarily apply to the work I have just mentioned. It would appear that there is a need for these different communities to make contact and become aware of their respective competencies. They have much to learn from each other. Work in CSCW has made the analysis of cooperative work in all its forms a central feature, and has produced a wealth of substantive literature on the practical accomplishment of human activities in work settings. This is an area which has been somewhat neglected in the cognitive ergonomic and engineering literature. de Montmollin agrees: "until recently models of collective activities were a rather neglected area in ergonomic research" and even when they were studied, the focus tended to be on ".....normative allocation of tasks, and to the corresponding design of prescribed communication, which is a different topic".

So, how might we go about opening up each other's communities of practice? There have been a number of attempts to synthesize different approaches, coming up with hybrid frameworks. One such framework has been elaborated by Shapiro. In a provocative and somewhat neglected paper from CSCW'94, Shapiro (34) attempts to walk a fine line between the traditions of ethnomethodology, psychology and participative design, in an effort to show how these different approaches and methods could possibly be mutually informing, and possibly be used together for certain design purposes. Taking the idea of satisficing from design, he argues against disciplinary Puritanism. He notes: "disciplines are the custodians of certain core perceptions which anyone setting out to achieve success in the design of certain kinds of system would ignore at their peril." As an example of the kinds of results one might be able to put together from such a frame, he mentions such points as:

"1. activities are socially organised and flexibly situated in context.
2. organizations. make deliberate strategic changes; these engage highly differentiated interest.
3. users can easily be alienated from a system for reasons of presentation, interface, and usability.
4. using a system imposes a variety of cognitive loads; these can be assessed only in relation to practice and training.
5. socio-technical systems are mutually constituting and adaptive.
6. users are the ultimate custodians of and experts in their own practices.
7. organizations and activities are continuously evolving.
8) the cost-benefit of systems should be optimised." (34)

While the extent to which it is possible to bring together concepts from possibly incommensurable conceptual viewpoints is questionable, there certainly is room for the further development and application of Shapiro's argument. There have been a number of recent attempts to bridge the gaps between different research communities. In Table 1 I list just a few events over the past few years, in which I personally have had some involvement, which were explicitly designed to encourage inter-disciplinary debate and discussion. The fruits of this collaboration should be visible across our communities in the next couple of years.

1992: ACM CSCW'92 Workshop on "Interdisciplinary Theory for CSCW" (Y.Waern, D. Shapiro)
1993: ACM InterCHI'93 Workshop on Rethinking Theoretical Frameworks for HCI (Y. Rogers, L. Bannon, G. Button)
1994: WEAAP Symposium: Opening up ATC Work: Behavioural, cognitive and sociological perspectives. (Bannon, L. & Shapiro, D.)
1997 : IEA Cognitive Ergonomics Symposium (E. Hollnagel)
European CSCW97 "FRINGE EVENT" Observation: Theory & Practice (L. Bannon, J. Hughes) Activity Theory, Distributed Cognition, Cognitive Ergonomics, Ethnomethodology
Revisiting the Allocation of Functions Issue: New Perspectives. Conference in Galway Ireland (Oct 97) (E. Fallon, L. Bannon, J. McCarthy)
EU COTCOS (Cooperation Technologies for Complex Work Settings) TMR Project. Report on conceptual approaches: Distributed Cognition, Activity Theory, Cognitive Ergonomics, Coordination Theory, Ethnomethodology, etc.
1998 : INRIA COOP'98 Conference May, Cannes, France. (HCI, DAI, Decision theory, CSCW, Cognitive Engineering, etc.)
EACE European Conference on Cognitive Ergonomics (Theme: collaboration in work) August 24-26 in Limerick (Programme Chair: E. Hollnagel, Conference Chair: L. Bannon)

Table 1: Creating Fora for Dialogue - a sampling

8. Concluding Remarks

In this paper I have argued that HF needs to develop an understanding of social ergonomics, as the next major frontier for the field. I have noted some of the reasons why I believe there is a shift in perspective from the cognitive to the social, and I have provided a glimpse of the work within the field of CSCW, which I believe has important insights into the social world for ergonomists. Possibilities for improved understanding of each tradition have been noted, ranging from attempts at hybrid conceptual frames to fora for debate about different approaches, their strengths and weaknesses for particular problems. While misunderstandings, wilful or naive, are inevitable in interdisciplinary communities, it is my view that despite the differences in disciplinary backgrounds and orientations, significant progress in mutual understanding has been made. It is important to note that what is being argued for here is not any simple *melange* of approaches, as it can be argued that fundamentally, they are not commensurate as they depend in many cases on different ways of viewing the world. That being the case, is not the attempt to "open up" our research communities a waste of time? I do not believe so. I believe that the tendency to build walls between people, including researchers, leads to an increasingly narrow and impoverished understanding of the world. This tendency is also evident in terms of what is acceptable, fundable etc. - i.e. it becomes institutionalised. While I accept that I am speaking "out of court" here, I do feel that there should be more openness on the part of major funding bodies in terms of accepting research proposals that utilize approaches other than the mainstream cognitivist position that has dominated most HF research in the past decade. Thank you for allowing me to address you, and I look forward to hearing about further attempts in the future at sensemaking across our disciplinary divides!!

9. References

1. Bannon, L. & Shapiro, D. (1995) Opening up ATC Work: Behavioural, cognitive and sociological perspectives. In N. McDonald , N. Johnston, R. Fuller, (Eds.) (1995) Applications of Psychology to the Aviation System, pp.216-222. Aldershort: Avebury Aviation.
2. Bannon, L. (1990) A Pilgrim's Progress: From Cognitive Science to Cooperative Design. AI & Society, 4,4, Fall Issue, 1990, 259-275.
3. Bannon, L. (1991) From Human Factors to Human Actors: The role of psychology and human-computer interaction studies in systems design. Book Chapter in Greenbaum, J. & Kyng, M. (Eds.) (1991) Design at Work.: Cooperative Design of Computer Systems. Hillsdale: Lawrence Erlbaum Associates, pp. 25-44.
4. Bannon, L. & Bødker, S. (1991) Beyond the Interface: Encountering Artifacts in Use. In J.M. Carroll (Ed.) (1991) Designing Interaction: Psychology at the Human-Computer Interface, pp.227-253. (New York: Cambridge University Press).
5. Carroll, J. (Ed.) (1991) Designing Interaction: Psychology at the Human-Computer Interface (New York: Cambridge University Press).
6. Thomas, J., & Kellogg, W. (1989). Minimizing Ecological Gaps in User Interface Design, IEEE Software, January, (pp. 78-86).

7. Hutchins, E. (1995) Cognition in the Wild. Cambridge, Mass.: MIT Press.
8. Winograd, T., & Flores C. F. (1986). Understanding computers and cognition: A new foundation for design. Norwood, NJ: Ablex.
9. Lave, J. (1988) Cognition in Practice. Cambridge: Cambridge University Press.
10. Vygotsky, L. (1978), Mind in Society. Cambridge, MA: Harvard University Press.
11. Leontiev, A. N. (1978). Activity, consciousness, and personality. Englewood Cliffs, NJ: Prentice-Hall.
12. Leontiev, A. N. (1981). The Problem of Activity in Psychology. In J. V. Wertsch (Ed.). The concept of activity in Soviet psychology. Armonk, NY: Sharpe
13. Engeström, Y. (1987). Learning by Expanding. Helsinki: Orienta-Konsultit.
14. Nardi, B. (Ed.) (1996). Context and Consciousness. Cambridge, Mass.: MIT Press.
15. Latour, B. (1996). Aramis, or the loveof technology. Cambridge, Mass: Harvard University Press.
16. Greenbaum, J. & Kyng, M. (Eds.) (1991) Design at Work.: Cooperative Design of Computer Systems. Hillsdale: Lawrence Erlbaum Associates.
17. Landauer, T. (1987). Psychology as a Mother of Invention. In J. Carroll, & P. Tanner (Eds.), Proceedings of CHI+GI'87, Human Factors in Computing Systems and Graphics Interface, Toronto, Canada, (pp. 333-335).
18. Landauer, T. (1987). Relations between Cognitive Psychology and Computer Systems Design. In J. Carroll (Ed.) , Interfacing Thought, Cognitive Aspects of Human-Computer Interaction (pp. 1-25). Cambridge, MA: MIT Press.
19. Landauer, T., 1991, Let's get real: A position paper on the role of cognitive psychology in the design of humanly useful and usable systems. In J.M. Carroll (Ed.) (1991) Designing Interaction: Psychology at the Human-Computer Interface, pp.60-73. (New York: Cambridge University Press)
20. Carroll, J. & Campbell, R. (1989) Artifacts as Psychological Theories: the case of human-computer interaction. (IBM RC 13454 (# 60225)), Yorktown Heights: IBM.
21. Whiteside, J., & Wixon, D. (1987). Discussion: Improving Human Computer Interaction - a Quest for Cognitive Science. In J. Carroll (Ed.), Interfacing Thought, Cognitive Aspects of Human-Computer Interaction. (pp 353-365). Cambridge, MA: MIT Press.
22. Weick, K. E. (1992). Sense making support systems. In D. Meadow, L.M. Jessup and J.S. Valecich (Eds.), Group Support Systems: New Perspectives, NY: MacMillan Press.
23. Strauss, Anselm. 1985. Work and the Division of Labor. The Sociological Quarterly 26 (1): 1-19.
24. Schmidt, K. & Bannon, L., 1992, Taking CSCW Seriously: Supporting Articulation Work, Computer Supported Cooperative Work (CSCW): An International Journal, 1(1-2), pp.7-40.
25. Anderson, R. (1994). Representations and Requirements: The value of ethnography in system design. Human-Computer Interaction, 9, 151-182.
26. Cooper,G., Hine,C., Rachel,J. & Woolgar, S. (1995). Ethnography and human-computer interaction. In P. Thomas (ed.) The Social and Interactional Dimensions of Human-Computer Interfaces. Cambridge: Cambridge University Press.
27. Hughes, J. & King, V. (1992). Sociology For Large Scale System Design. Paper For Software & Systems Practice: Social Science Perspectives, CRICT Conference, Reading, Nov 30th - Dec 1st, 1992.
28. Hughes, J. A., D. Randall & D. Shapiro. (1992) 'Faltering from Ethnography to Design', in Proceedings of "CSCW '92: Sharing Perspectives"; the Fourth International Conference on Computer Supported Cooperative Work, Toronto, 2-4 November 1992, pp 115-122.
29. Rognin, L. (1998) Working in Control Rooms. Cooperation and Communication in Shared Workspaces. Proceedings of COOP'98, May 27-29, Cannes, France
30. Rognin, L. & Bannon, L. (1998) CAT: Investigating Cooperation and Communication in Air-Traffic Control. Ensuring safety through collaborative work arrangements. The case of Shannon Air-Traffic Control. IDC internal report. June 1998

31. Rognin, L., Salembier P. & Zouinar M. (1998). Cooperation, interactions and socio-technical reliability: the case of Air-Traffic Control. Comparing French and Irish settings. Proceedings of European Conference on Cognitive Ergonomics, ECCE-9, August 24 - 26,Limerick, Ireland.

32. Bowers, J. (1997) Allocation of Function: A CSCW Perspective. Keynote Address, Conference on Revisiting the Allocation of Functions Issue: New Perspectives, Oct. 1-3, Galway, Ireland,1997.

33. de Montmollin, M., 1991, Analysis and Models of Operator's Activities in Complex Natural Life Environments. In J. Rasmussen, H. Andersen, & N. Bernsen (Eds.) Human-Computer Interaction, Research Directions in Cognitive Science: European Perspectives, vol.3, (Hove: Lawrence Erlbaum).

34. Shapiro, D. (1994). The Limits of Ethnography: Combining Social Sciences for CSCW. Proceedings CSCW'94, Raleigh, North Carolina, USA. pp.417-428.

Measurements And Models, Models And Measurements: You Can't Have One Without The Other

Erik Hollnagel, Ph.D.
Professor, Graduate School of Human-Machine Interaction
University of Linköping, S-581 83 Linköping, Sweden
eriho@ikp.liu.se

1. SUMMARY

The paper discusses the relation between measurements and models. Two conditions are identified: one where measurements refer to articulated models, and one where measurements refer to folk models. It is argued that measurements should refer to the performance characteristics of the joint system, rather than of assumed cognitive functions. Furthermore, that it is not meaningful to describe described independently of the context. A proposed measure relates to the orderliness of performance, i.e., the level of control that the joint system has over the situation. The possible details of this measure are outlined.

2. MEASUREMENTS MUST HAVE MEANING

In the study of human performance the definition or specification of what one should measure is undoubtedly the most important problem, whether for individual or crew performance. Measurements must meet three essential requirements: (1) they must be **possible**; (2) they must be **reliable**; and (3) they must be **meaningful** or **valid**. Very few of the measurements that are used in practice meet all three requirements. All, of course, meet the first requirement - although it is quite possible to define hypothetical measurements that cannot be made for reasons of either philosophy or technology. Some measurements meet the second requirement, and fewer the third.

One important distinction is whether a measurement is theory driven or theory begging. A **theory driven** measurement is derived from an articulated model of a phenomenon or a functional relationship, in the sense that the semantics - or the meaning - of the measurement is provided by the model. A **theory begging** measurement is derived from an indistinct or incomplete model, often referred to as a folk model. The meaning of that is a commonly held idea or notion, often shared among experts and non-experts, about the nature of an everyday phenomenon. Folk models are very common within psychology and the behavioural sciences, probably because we all have "privileged knowledge" about how the mind works (Morick, 1971). Folk models are not necessarily incorrect, but compared to articulated models they are incomplete and focus on descriptions rather than explanations. In contrast to articulated models they are also very difficult to prove wrong. The distinction between articulated and folk models has consequences for the types of performance measures that can be used.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

2.1 Measurements Based On Articulated Models

The definition of a measurement depends on how the domain from which the measurement is taken is thought of or conceptualised, even if this is done implicitly rather than explicitly. (The domain is usually referred to as the target system, which in the behavioural and cognitive sciences denotes the real-life phenomenon for which a measurement is sought.) The definition of a measurement therefore presupposes a clarification of what the model behind the measurement is, and how the target system can be adequately described. A model is by definition a simplified representation of the salient features of the target system. The model constrains what can be measured by describing what is essential performance and the model parameters thereby become the basis for specifying the measurements. Since it is impossible for the model to contain all the parameters of the target system, the characteristics of the model define the important measurements.

Most models are of the structural type, i.e., they represent the functions of a system (in particular, of a human) by means of some hypothetical structures or elements of mental machinery, as well as by the relations between them. A good illustration of that is the conventional information processing model, of which an exemplar is shown in Figure 1. Here human actions are described as emanating from a relative simple system of functional units, such as a number of stores (sensory store, working memory, long term memory), a decision making unit, an attention regulating unit, etc. This description implies that measurements should be related to the theoretically defined functioning of these units, as well as to the links (or information channels) between them.

In the 1960s and 1970s the modelling efforts focused on the fundamental information processes, particularly those related to perception and memory (Attneave, 1959; Lindsay & Norman, 1977). Measures were defined according to the models, as for instance limited capacity central processing or levels of processing in multi-store memory models (cf. Norman, 1976). The details of the models, and the constrained character of the phenomena being studied, allowed very specific measurements to be proposed. Later on, when the interest turned from the mechanisms of perception and memory to the cognitive functions that were part of e.g. problem solving or reasoning, it became more difficult to propose theory based measurements. Instead data were found through such means as verbal protocols and introspective accounts. The

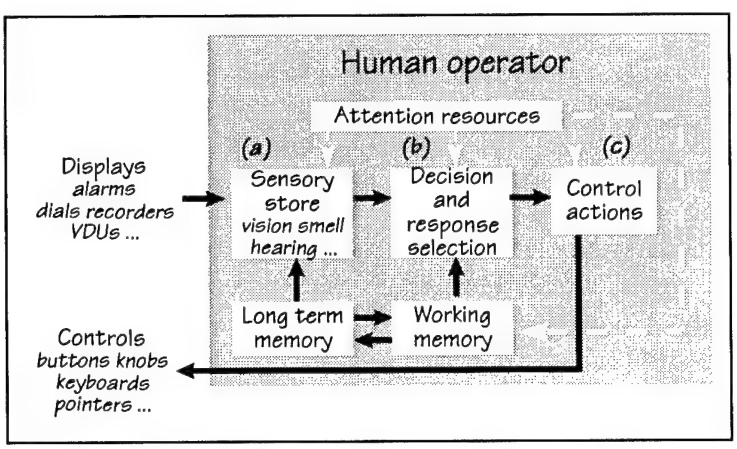


Figure 1: A characteristic information processing model.

models started to look outward to how people interacted with the environment, although still mostly as part of contrived tasks, and the measurements reflected this change. In cases where the research went out of the laboratory, or at least looked at problems taken from real work situations, the focus turned to general traits such as attention, workload, etc.

2.2 Measurements Based On Folk Models

If an articulated theory is not available, the measurements can be derived from a general understanding of the characteristics of the system and of the conditions of human work. An example of that is workload and, more recently, situation awareness (Endsley, 1995). Workload, for instance, reflects the subjective experience of mental effort, which is so pervasive that it can be applied to practically all specific situations. Furthermore, workload is acknowledged to be an important causal factor in the folk models of human performance, i.e., the accounts of causal explanations of events, typically accidents or incidents. It is thus a measure that is defined by consensus, rather than by reference to a model. In fact, the models have usually come afterwards.

Folk models describe measures that reflect an important aspect of the operators' situation, but usually related to intermediate "cognitive" states rather than to the actual performance. It is assumed that the measurement is a valid substitute for actual performance measurement, because refers to an essential intermediate or intervening state. It is also assumed that the measurement is affected by the performance conditions to the same extent and in the same manner as the actual performance. These assumptions are illustrated by Figure 2.

In relation to operator performance, measurements proposed by folk models represent the commonly held notions about the nature of human work, and specifically about the nature of human cognition. At present, i.e., in the mid-1990s, the main concepts that are used to describe the cognitive aspects of work are, for instance, attention control, working memory management, mental workload, situation awareness, the operator's mental model, the processes or patterns of reasoning, and meta-cognitive self-monitoring. It is clearly easier to propose a measurement for some of these concepts than for others, although the ease by which measurement tools can be developed do not necessarily reflect the significance or validity of that measurement.

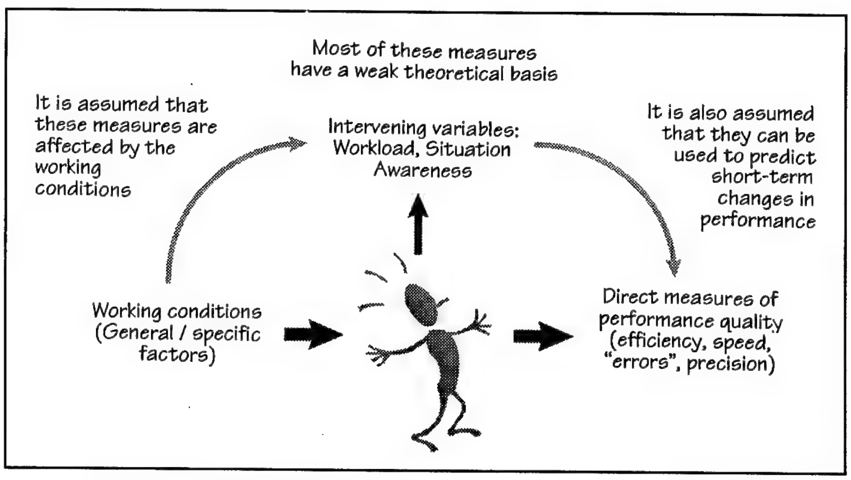


Figure 2: Loosely defined measurements.

2.3 Measurement Possibility vs. Interpretation

If we consider the range of measurements that are typically employed in empirical research, and particularly in experimental (laboratory) research, it is possible to discern a relation between how easy it is to make a measurement and how meaningful it is. Figure 3 shows this relationship for some of the more common measurements.

As Figure 3 suggests, the various measurements seem to be distributed about a diagonal. Many measurements are thus relatively easy to make, but have a limited theoretical basis and are difficult to interpret. This is typically the case of measurements can be easily recorded by mechanical means such as measurements of physiological variables (heart rate) or overt performance (audio, video recordings). Other measurements have an acceptable theoretical foundation, but are either difficult to make or difficult (and laborious) to interpret. Examples of that are eye movement recordings or performance "errors". It would clearly be very useful if measures could be proposed which were both easy (and reliable) to make and meaningful. It follows from the preceding arguments that such measures must be based on an articulated model, rather than a folk model.

3. COGNITIVE SYSTEMS ENGINEERING

There is clearly a need to propose measures that are both meaningful and relatively easy to make - and to the extent that these demands are not both achievable, it is more important to have measures that are meaningful. This requires a basis that on the one hand is more articulated than folk psychology, and on the other better linked to the requisite variety of human performance than information processing models. In the remaining parts of the paper I shall to to describe how cognitive systems engineering (Hollnagel & Woods, 1983) can be used as such a basis.

3.1 What Is A Cognitive System?

One of the motivations for the development of CSE was the need to provide a common set of terms by means of which the interaction between people and machines could be described. The notion of information processing was widely used, but

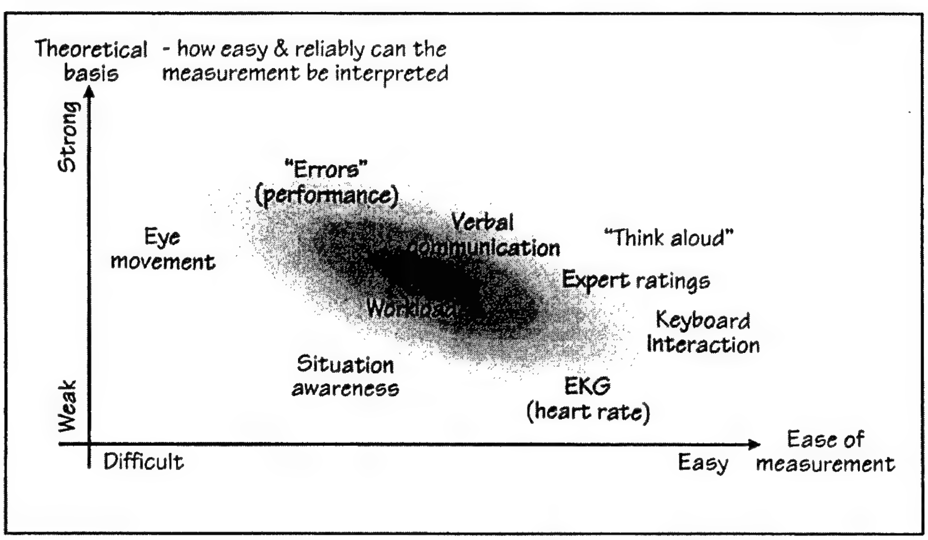


Figure 3: Meaning of measurements.

actually implies a strong technological metaphor which imposes serious limitations on both models and measurements. The solution was to focus instead on the essential characteristics of efficient performance, namely that it is directed towards a goal, that it makes use of past experience, and that it adapts to the current conditions. Systems that show these traits were called cognitive systems. People, obviously, are natural cognitive systems, while machines in many cases can be considered as artificial cognitive systems.

An important premise for CSE is that **all work is cognitive**. There is therefore no need to distinguish between cognitive work and non-cognitive work or to restrict cognitive work to mean the use of knowledge "to intentionally realise the possibilities in a particular domain to achieve goals". Everything we do requires the use of cognition with the possible exception of functions regulated by the autonomic nervous system. The cognitive content of even basic skills becomes obvious as soon as we try to unpack them or apply them under unusual circumstances, such as walking down a staircase in total darkness. The fact that we habitually are able to do a great many things without thinking about them or paying (much) attention to them does not make them non-cognitive. Similarly, CSE considers the use of tools without making a distinction between cognitive tools and non-cognitive tools. A tool, such as a bicycle, may have been developed to support a predominantly manual function but anyone who has ever tried to teach a child to ride a bicycle will be keenly aware that this involves a very high level of cognition.

Formally, a cognitive system is characterised as being able to modify its pattern of behaviour on the basis of past experience to achieve specific anti-entropic ends. This is done by using information about itself and the situation, where the information can be prior information (knowledge, competence), situation specific information (feedback, indicators) and constructs (hypotheses, assumptions). The control can be complete or partial and depends to a considerable extent on the ratio of unexpected information to expected information. Cognitive systems engineering is concerned with the analysis and design of **joint cognitive systems**, i.e., cognitive systems that co-operate to achieve their ends. Part of this concern is to define appropriate measurements for the performance of the joint system.

3.2 Measurements According To CSE

According to the models of CSE, a critical aspect of maintaining control is the ability to sustain a proper balance between feedforward and feedback. This in turn depends on the ability of the cognitive system, in particular of the operator, correctly to understand how the system works and anticipate how it will respond to specific actions. (Note that in this description the system may comprise both the technological system or the process, other people who are present in the situation and who take part in the work, or both together.) Situation understanding and anticipation are, however, both notably difficult to measure. It is not obvious how appropriate indicators can be found, nor how they should be interpreted.

One basic premise for proposing a measurement is that **cognition cannot be measured independently of the context**, whether it is considered in terms of cognitive states or cognitive functions. Measurements must refer to the performance of the joint system, rather than to the performance of either of the component systems. This mirrors the objective of a cognitive task analysis which is to reveal the demands to the joint cognitive system whereas the

objective of a classical task analysis is to reveal the demands to human cognition. The issue in both cognitive task analysis and performance measurements is how well the Joint Cognitive System can accomplish the required work rather than whether the human can meet the demands of the machines.

Since cognition cannot be measured independently of the context, one category of measurements refer to working conditions or the context. One suggestion for that are the Common Performance Conditions (Hollnagel, 1998a), or more generally the category of performance shaping factors (Swain, 1989). The performance shaping factors are, however, not independent variables in the sense that they can be controlled or varied at will, or in the sense that they are independent of each other. On the contrary, they are clearly coupled and it is therefore necessary to develop some kind of model which can account for the coupling.

Another category of measurements refers to the concept of control, which can be specified further in terms of maintaining control and losing control. The relations between control and performance are described by the Contextual Control Model (COCOM), which has been developed from the principles of CSE (Hollnagel, 1998b), and which therefore provides the basis for interpreting proposed measurements. The level of control is, however, a hypothetical state which cannot be measured directly; it expresses itself in terms of the orderliness of performance for which some measures can be proposed.

The purpose of COCOM is to offer a way of describing human behaviour, both in terms of developing causal explanations for something that has happened and in terms of enabling predictions of what may happen in the future. One important part of COCOM is the assumption about the characteristic types of performance that are associated with assumed control mode. In addition, the model also describes how performance is determined by the common performance conditions. The model makes a distinction among four characteristic control modes (Hollnagel, 1998a):

- In **scrambled control** the choice of next action is in practice unpredictable or haphazard. Scrambled control characterises a situation where there is little or no thinking involved in choosing what to do. This is typically the case when the task demands are very high, when the situation is unfamiliar and changes in unexpected ways, when thinking is paralysed and there accordingly is a complete loss of situation awareness. The extreme case of scrambled control is the state of momentary panic.

- In **opportunistic control** the next action is determined by the salient features of the current context rather than on more stable intentions or goals. The person does very little planning or anticipation, perhaps because the context is not clearly understood or because time is too constrained. In these situations the person will often be driven either by the perceptually dominant features of the interface or by those which due to experience or habit are the most frequently used, corresponding to the similarity matching and frequency gambling heuristics described by Reason (1990). The result is often functional fixation (De Keyser, et al., 1988).

- In **tactical control** performance is based on planning, hence more or less follows a known procedure or rule. The planning is, however, of limited scope and the needs taken into account may sometimes be *ad hoc*. If the plan is a frequently used one, performance corresponding to tactical control may seem as if it was based on a procedural prototype - corresponding to

e.g. rule-based behaviour. Yet the regularity is due to the similarity of the context or performance conditions, rather than to the inherent "nature" of performance.

- In **strategic control** the person considers the global context, thus using a wider time horizon and looking ahead at higher level goals. The strategic mode provides a more efficient and robust performance, and may therefore seem the ideal to strive for. The attainment of strategic control is obviously influenced by the knowledge and skills of the person, i.e., the level of competence. In the strategic control mode the functional dependencies between task steps (pre-conditions) assume importance as they are taken into account in planning.

3.3 The Orderliness Of Performance

COCOM can be used to propose performance measurements that correspond to the control mode and which can be used to characterise the orderliness or the degree of organisation and structure in the person's performance. One possible measure is the length of time a person spends in following a specific goal, i.e., whether there are frequent shifts between goals, whether the same goal is pursued for some time, whether the shifts are regular or irregular, i.e., whether an underlying principle can be recognised. According to COCOM the pattern or profile for the opportunistic control mode would be many shifts between goals and therefore a limited time spent on each goal, whereas the pattern for tactical (attended) control mode would be few shifts between goals and therefore a longer time spent on each goal, relative to the time constant of the process. Another measure is the way in which a person allocates resources (for instance, time) among multiple goals - specifically whether it is possible to infer a clear strategy in that, or whether it seems to be of a more random nature. In COCOM terms, performance in the tactical (attended) control mode will be better than performance in the opportunistic control mode, referring to typical criteria such as number of mistakes and errors, efficiency, safety, etc. Yet another type of measure is how people respond to information from other sources (the process or other people), for instance whether the response occurs within a reasonable time (time window), too late, or not at all, and how adequate the response is (adequate and complete, adequate but simplified, inadequate or irrelevant, missing, cf. Miller, 1960).

Table 1: Proposed measurements of orderliness of performance.

COCOM control mode	Orderliness	Activity types	Responses to input information	Timing of responses
Strategic	Well ordered, smooth	Monitoring, scheduling	Appropriate, anticipatory	Precise
Tactical (attended)	Ordered, efficient	Control, regulating	Appropriate	Timely
Tactical (unattended)	Ordered, but detached	Feedforward driven	Appropriate, but inconsistent	Possibly delayed
Opportunistic	Vaguely ordered	Feedback driven	Simplified	Lagging
Scrambled	Disordered	Action (observation)	Unpredictable	Unpredictable

The proposed measures focus on how orderly or well-organised performance is, rather than on whether performance is correct or incorrect relative to some criterion. There are two reason for being interested in the orderliness of performance: firstly, because the orderliness of performance depends on the general

performance conditions; and secondly because orderly performance is better than disorderly performance, in the sense that the reliability and efficiency is higher. A summary of the proposed measures is provided in .

It is quite possible to analyse the performance in terms of the proportion of specific categories, for instance observations and actions, and also in terms of whether the actions represent few or many simultaneous goals or lines of activity. Although analyses necessarily are qualitative, they can be performed quite rigorously and methodically, e.g. Hollnagel et al., 1981, Sanderson & Fisher (1994). At the moment, the greatest problem is the need to interpret the measurements afterwards. However, because the measurements are based on an articulated model, the interpretation can be approached in a systematic manner thereby providing a high level of intersubjective agreement.

4. REFERENCES

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods, and results.* New York: Holt, Rinehart & Winston.

De Keyser, V., Masson, M., Van Daele A. & Woods, D. D. (1988). *Fixation errors in dynamic anc omplex systems: Descriptive forms, psychological mechanisms, potential countermeasures.* Liege, Belgium: Universite de Liege, Working Paper Series No. 1988-014.

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32-64.

Hollnagel, E., Pedersen, O. M., & Rasmussen, J. (1981). *Notes on human performance analysis* (Risø-M-2285). Roskilde, Denmark: Risø National Laboratory, Electronics Department.

Hollnagel, E. (1998a). *Cognitive reliability and error analysis method – CREAM.* Oxford: Elsevier Science.

Hollnagel, E. (1998b). Context, cognition, and control. In Y. Waern, (Ed.). *Co-operation in process management - Cognition and information technology.* London: Taylor & Francis

Hollnagel, E. & Woods, D. D. (1983). Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies*, *18*, 583-600.

Lindsay, P. H. & Norman, D. A. (1977). *Human information processing: An introduction to psychology* (2nd Ed.). New York: Academic Press.

Miller, J. G. (1960). Information input overload and psychopathology. *American Journal of Psychiatry*, *116*, 695-704.

Morick, H. (1971). Cartesian privilege and the strictly mental. *Philosophy & Phenomenological Research*, *31(4)*, 546-551.

Norman, D. A. (1976). *Memory and attention* (2nd. ed.). New York: Wiley.

Reason, J. T. (1990). *Human error.* Cambridge, U.K.: Cambridge University Press.

Sanderson, P.M., & Fisher, C. (1994). Exploratory Sequential Data Analysis: Foundations. *Human-Computer Interaction*, 9, 251-317.

Swain, A. D. (1989). *Comparative evaluation of methods for human reliability analysis* (GRS-71). Garching, FRG: Gesellschaft für Reaktorsicherheit.

Cognitive Engineering: The Latest Fad or a True Step Forward as an Approach to Complex Multi-Person System Analysis and Design?

Robert G. Eggleston
Air Force Research Laboratory
Human Effectiveness Directorate
2255 High Street
Wright-Patterson, AFB, OH 45433-7022
USA

1. SUMMARY

Cognitive Engineering is a user-centered approach to the analysis and design of complex systems. It is especially applicable to adaptive systems such as those involving teams of people and/or intelligent machine agents. To date, Cognitive Engineering has at best made limited penetration into design practices for military systems. To be sure, cognitive issues are of great concern in military system design. It is less clear, however, what effect Cognitive Engineering is having on design practice. This paper looks at the gap between Cognitive Engineering in the research community and its use in the crew station design process for military systems. Five characteristics of Cognitive Engineering are identified and contrasted with prevailing design practice. These factors illustrate conceptual and practical differences with current crew vehicle interface (CVI) practices. Other factors that tend to limit understanding of Cognitive Engineering by system mangers and design practitioners are also mentioned. In addition, five weaknesses in current practice are identified. It appears at least some of the weak areas could benefit from the use of Cognitive Engineering. However, there are also open issues concerning the ability of Cognitive Engineering to handle certain characteristics of military system design challenges. These are discussed, as well. The likelihood that Cognitive Engineering will be integrated into military systems engineering and crew station design practices may well depend largely on its ability to produce design artifacts that help software engineers design aiding subsystems interactively with the crew vehicle interface (CVI). Ideally, these separate design activities would be more tightly coupled, and Cognitive Engineering offers an approach to achieve such a coupling.

2. INTRODUCTION

Cognitive Engineering is a theoretically based design framework and an engineering practice whose aim is to produce robust, cooperative systems that aid human performance. It is particularly relevant for the design of complex adaptive systems to be used in high demand work situations. (Such work environments are exemplified by conflicting goals, changing uncertainties, a large resource pool, multiple co-occurring tasks and events, high stress and high workload.) High-demand work is a common characteristic of military operations. Cognitive Engineering is an example of a user-centered design approach to the development of human-machine systems for use in such environments. This framework considers design issues that emerge from an analysis of individual task work, team work, and collaborative work in the course of solving work problems to accomplish system goals.

Cognitive Engineering is a relatively young design technology. Initial presentation of cognitive engineering ideas and concepts first appeared in the early 1980's in sources generally available to the research community. However, it has only been in the past few years that it has gained notice and attention from a wider audience of human factors practitioners and system developers. Cognitive Engineering (CE) is viewed by many in the research community as a critical technology advancement that is essential to the development of complex, modern day systems that incorporate a significant level of machine intelligence. In spite of the promise of Cognitive Engineering for military system design, there still remains, as Woods et al. recently put it, a "gap between user-centered intentions and actual design practice" (1996, pg 967). They suggest that development deadlines and other time pressures of the design process for large-scale system development tend to limit the type of depth of analysis devoted to problem understanding required by CE. They believe this stems from an organizational problem- a lack of commitment to balance investment across understanding, innovation, and refinement over multiple projects. While this organizational

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

factor may make it more difficult to practice CE in a large-scale development effort, several other aspects of current design practices also contribute to the gap between user-centered intentions and practice.

This paper looks at the gap between cognitive engineering in the research community and its use in the crew station design process for military systems. It has been my experience that many engineers involved in crew station design, from the earliest stages of concept exploration to detail design for military systems, have had difficulty recognizing differences between Cognitive Engineering practices and current design engineering practices. Program managers also often seem unclear about what Cognitive Engineering offers, how it is accomplished, what products it produces, and how it can be integrated into the work structure for system development. The more skeptical practitioners and managers may even wonder privately: Is CE just the latest fad? Is it like a "new" management system that sounds good, offers a few new insights, but when pressed by hard problems quickly fades because it is no better at handling them than currently used methods? Or is CE a genuine step forward that will add lasting value to military system design and development processes and products?

The goal of this paper is to inspect CE in an effort to better reveal how it is different from current crew station design practices. It can provide a starting point for a conceptual analysis of design practices that may help managers and practitioners decide if it is worth investing in a more thorough study of CE for consideration for use in analysis and design.

The paper begins with a brief overview of the major system acquisition and development process, with special attention given to the subprocess of crew station design. Next, it attempts to distinguish conceptual differences between CE and features of current design practices. Five characteristics are used to illustrate the sometimes subtle differences between prevailing crew station design practices and Cognitive Engineering practices. The paper concludes by pointing out some weakness in current practice and indicates some areas where further development is needed in CE. No attempt is made to provide a comprehensive review of CE. The topic is complex and beyond the scope of this paper. Excellent coverage of CE can be found in Norman, 1984, 1987; Hollnegal and Woods, 1983; Roth and Mumaw, 1992; Woods and Roth, 1988a, 1988b. Rasmussen, 1983, 1986, and Rasmussen, Pejtersen, and Goodstein, 1994).

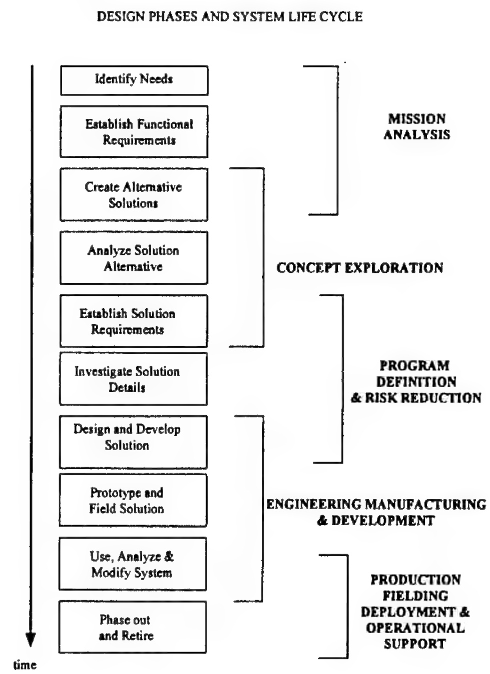


Figure 1. System Acquisition and Design. Adopted from Rouse and Cody, 1988.

3. CREW STATION DESIGN PROCESS

Crew station design is accomplished within a major systems acquisition process. A simplified view of the total acquisition process is depicted in Figure 1. Acquisition, and the attendant research and development, is accomplished in a top-down manner, beginning with a statement of needs. If currently fielded systems are unable to meet the need, then planning for a new system begins. The ten steps shown in the process cover development from inception through the fielding of a system, including modifications and eventual retirement. These steps are separated into five broad acquisition phases, as shown. Passage from phase to phase is contingent upon satisfying senior decision makers based on a comprehensive review of the program. Historically, it has taken large scale weapon systems programs about 13-15 years to reach the production phase.

A simplified view of the crew station design process is depicted in Figure 2 in relation to the procurement process. Early activities in this process are centered around expanding the initial mission description as a means of increasing

understanding of mission characteristics that, in turn, stimulate the need for system requirements. This expanded mission model, identified as level 2 b (L2b) in Figure 2, is used to support initial workload assessments and other preliminary analyses that are used to allocate functional responsibility to the crew, the machine, or to both (i.e. some form of shared responsibility). In parallel with these activities, crew station designers evaluate crew technologies and accomplish various trade studies. Trade studies are used to select among alternative means for accomplishing a function or an aspect of a function, and also to prioritize technologies for inclusion in the system. Typically there are many different ways to satisfy a mission function or requirement. Usually a development will only include a subset of all of the technologies proposed by each functional area leader based on a judgment of what is best for the system as a whole, given an assessment of mission priorities and development factors such as cost, schedule, and technology risk. As more knowledge is acquired, crew functions are expressed with more details in terms of task sequences that are further analyzed for work demand and information requirements to accomplish the work successfully. The outcome of the set of analysis activities serve as major input into the design synthesis activities that eventually culminate in the crew vehicle interface. Other functional aspects of the system, of course, are being designed in parallel with the crew station which also feed into the CVI design.

The design process is complicated by the fact that mission concepts, functional priorities, and technology options change throughout the course of the design activities. These and other factors contribute to the fact that the synthesized design is not uniquely determine by the preceding analyses. Rather, the synthesized design emerges from the analysis, experience, creativity, and persuasive abilities of the members of the design engineering staff.

This description of a prototypical crew station design process omits many details, including design iterations and tests to discover problems and gauge progress.

Crew station design is by nature user-centered, since the issue is to establish a linkage between the user and the engineered system to be used in work. Yet, the design engineer may approach the problem in a machine-centered way: i.e. make engineering decisions that attempt to optimize hardware or software factors even if they make it more difficult for the user to operator the system. Cognitive Engineering has been advocated as a user-centered approach to system design because it treats interface design from the perspective of being a support system for the user that facilitates system use in solving the work problem. Given this emphasis on use, some advocates of cognitive engineering have referred to this as use-centered design (Flach and Dominguez, 1995). This support system perspective begs the question: Does CE differ from prevailing machine-centered or more traditional user-centered design practices? If so, in what ways?

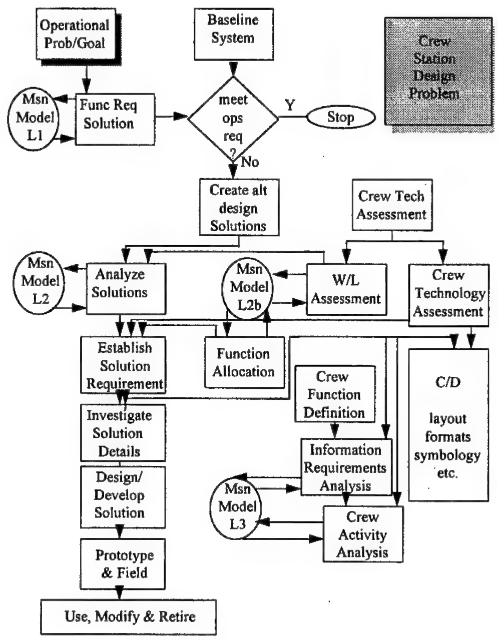


Figure 2. Nominal crew station design process.

4. WHAT IS DIFFERENT?

CE offers principles for interface design and embraces the use of a diverse set of analyses of a work domain and the work force. However, it is a general framework that can be implemented in different ways. As a result, specific practices may vary. Nevertheless, there are general aspects of CE that guide design practices, and some of these are different, at least conceptually, from prevailing crew station design practice. I have identified five characteristics of CE that in broad terms reflect the shift in thinking that motivates differences in practices when this framework is used. Figure 3 presents these five factors and indicates how they differ from current design practices. A discussion of each factor follows.

Current Practice	Cognitive Engineering
• Design crew station and crew-aiding functions as separate but related subsystems	• Design a user support system as an integration of the crew station and crew-aiding functions
• Use behavior task trajectories to guide interface design	• Design a workspace using constraint boundaries and work strategies as guides
• Analyze cognition as a stage of information processing	• Analyze cognition as internal and external, distributed, shared, and public
• Analyze the mission in terms of function allocation, task activities, and information requirements	• Analyze the mission in terms of problem solving, naturalistic decision making, error induction, and expertise
• One-dimensional functional decomposition used for analysis and design	• Two-dimensional decomposition used for workspace analysis and design

Figure 3. Five characteristics that distinguish CE from current military system design practice.

4.1 Integrated support system. Perhaps the single most important difference is that cognitive engineering begins with the viewpoint that the crew vehicle interface (CVI) is a *user support system.* Historically, the CVI has been viewed in terms of displays and controls, with the actor viewed as an operator. From this perspective, the design problem is to develop suitable displays and controls to allow the actor to operate the system effectively and efficiently. This orientation still prevails today for the design of CVIs for advanced military systems.

The cognitive engineering view regards the actor first and foremost as a *problem solver* and that the *crew station is provided to support the actor solve problems of work as they arise.* The problem solving viewpoint accentuates the fact that in today's systems a great deal of actor work involves abstract thinking about the conditions and job at hand and thus places heavy demands on cognitive activity that is essential to job success. When the crew station is viewed as a user support system, then the design task changes from one of only display-control design to one of integrated design of displays, controls, actor-agent transactions, and agent behavior as they relate to problem solving. In other words, design of the crew stations interfaces also includes the design of machine automation (aiding subsystems). These two aspects of the of the CVI design problem need to be designed interactively and jointly as a connected whole.

In contrast to this integrated user support system view, current design practice separates responsibilities for the design of the user interface from the design of aiding automation. The design group responsible for the traditional "human factors" aspects of the crew vehicle interface (e.g. display formats, symbology, layout, etc.) is different from the team responsible for the design of aiding technology to be embedded in the system. Indeed, there is considerable interaction between these two groups, but they have entirely different responsibilities. The computer engineer, i.e. the artificial intelligence specialist, is responsible for developing aiding behavior, to meet functional expectations established for the system. The crew station engineer is responsible for the display/control details of actor interaction with the system. Currently the design products produced by the crew station designers used to meet their responsibilities are in general of little use to the designer of the aiding subsystem, and vice versa. As a result, after all elements of the system have been integrated into a final design solution, it is not uncommon to discover the planned aiding is brittle in unintended ways and thus does not always fit in well with the flow of work demanded by the mission situation.

4.2 Workspace vs. work trajectories. The aim of cognitive engineering is the development of *flexible* user support systems that continue to provide support under unusual and abnormal problem solving situations as well as prototypical ones. Flexible support by the user interface, of course, is also a long standing goal of human factors engineering. Classical human factors and cognitive engineering differ in terms of the prototypical approach taken to achieve this goal. The traditional approach involves an analysis of individual behavioral trajectories that an actor takes to accomplish a task or set of tasks. Activities required to complete the "full set" of tasks under a range of conditions are used to guide the design. Thus the design is constructed to explicitly support the sum of the expected behavioral trajectories. In effect, these behavioral sequences are used to define the actor's workspace. Flexibility is achieved with this approach in two main ways: (1) multiple interaction methods are provided that permit the same action to be completed (e.g. voice, cursor, key input paths); and (2) a small set of common input methods are used to accomplish a large range of system actions (e.g. use the same cursor slewing technique to change waypoints, select menus, designate targets, etc.).

CE approaches the design problem in an entirely different way. According to this design theory, the workspace is established not by behavior trajectory analysis, but by an analysis of the factors that serve to establish constraint boundaries, plus an analysis of work strategies

that *can* be followed to solve work problems, given different constraint profiles. The aim is to establish a workspace that can easily reveal the current situation and its underlying dynamics and readily support action taking as deemed necessary based on the problem solving skills (expertise) of the actor. In this way the workspace for the actor is made isomorphic with the problem space confronted at the moment.

The trick in this design approach is to be able to identify the constraints that influence the dynamics of work and how it is understood by skilled practitioners. This presents a difficult analysis problem, because constraints may be derived from many different sources: some physical (e.g. laws of physics), some based on system resource limits (e.g. sensors), and some based on organizational factors (e.g. rules of engagement). Various methods are used to gain understanding about the workspace and the conditions for various degrees and types of breakdown conditions. A wide range of factors are used to set the final specifications for the workspace and once this is accomplished then support for a wide range of work strategies (e.g. skilled based, rule based, knowledge based) can be planned to provide flexible interface methods (e.g., see ecological interface design theory Rasmussen and Vicente, 1992, Rasmussen et al., 1994) for a more detailed explanation of this view).

4.3 Cognition. Another conceptual difference that is implicated in crew station design practices revolves around the very notion of what is meant by the term cognition. In current practice, cognitive behavior is mainly analyzed as a property of an individual actor. This stems from an information processing view of human behavior. Cognition is a stage in information processing that occurs in the head. It follows, then, that a cognitive task analysis used in system design involves assessing the demand placed on this stage of information processing. Prevailing design practice has included more in-the-head cognitive factors in task analysis and other aspects of CVI design in recent years. This trend began in the early 1980's after concerns over mental workload gained attention in military systems design (McCracken and Aldrich ,1984, Alrdich et al, 1988; and North, 1988).

Many, but not all, in the CE community emphasize a different aspect of cognition. They focus on cognition as distributed and shared across people and artifacts in the environment (see,e.g. Hutchins, 1995). Cognition is involved in social interaction, coordination, cooperation, collaboration and other aspects of the environment. That is, cognition is *in* the environment or a property of the environment. In other words, knowledge exists in the world relative to an actor. The simple act of seeing a post-it, for example, can serve to remind an actor that a task is due or an appointment is coming up. The actor does not have to read the text; it is enough to just notice the note for the cognition to form, pop up or be instantly recognized. This view of cognition as distributed and in the environment essentially treats it as an affordance that can simply be picked up, once it is in place based on experience and prior events. It is easy to see how social interactions can carry cognition. Seeing how a person acts often informs us about their intentions and may support predictions of future events. Cognition then is not only in the head, but it can be externalized and simply picked up from the environment in the course of working. CE is sensitive to this aspect of cognition. As a result, it tends to analyze work situations in these terms and attempts to use this knowledge in the design of the actor's workspace. It changes the nature of information requirements analysis, since information often is emergent and based on prior acts by a performer.

4.4 Mission and work analysis. Mission and task analyses are important aspects of CVI design. The very process of forming a mission description serves to educate the CVI design team about the problem domain and adds clarity to the design task. Mission and task analysis is evolutionary, constantly under going change, as needed, to support the next level of design detail, and to stay current with the changing view of potential courses of action a potential adversary could take (i.e. changes in the job problem). Analysis models are constructed and used by many groups thoughout system design, including the crew station design group. At the same time, a mission (usually a composite of some type containing the most difficult features of several missions) is sanctioned and certified for use for the entire system development to provide a common framework.

Mission descriptions and work analysis are also critical elements in CE based design practice. Differences exists between these two practices, however, in terms of some of the things they look for and the analysis products they produce. In crew station design, mission analysis follows a linear progression of activities that usually

begins with a workload analysis, followed, in turn, by functional requirements analysis , function/task allocation analysis, and information requirements analysis. The approach to understanding is inherently reductionistic: Start with a broad view of the work problem and reduce it to behavioral trajectories and information requirements. This, then, become the elements that drive the design.

In contrast, CE concentrates on a set of activities that help the analyst to better understand mission work in terms of actor problem solving and naturalistic decision making, placing special emphasis on discovery of difficult situations, hard problems, and situational factors that tend to induce errors in performance. The goal is to understand how experts perceive the mission situation, how they solve emerging problems, and the set of factors that together invite errors. Further, an attempt is also made to discover the fundamental properties of the work problem, those that are independent of the tools and systems used to solve it. Sometimes this is referred to as understanding the semantics of the domain. All of these factors, of course, are of interest to the crew station designer, but they usually do not receive the level of analytic attention devoted to them by CE. Rather, the current approach tends to lock-on to local elements of the mission scenario and aims to identify a solution that is hoped to be flexible enough to handle mission variations. It does not attempt to model the problem at multiple levels of abstraction that make clear the full scope of human problem solving.

These differences perhaps stem from the fact that classical design attempts to concentrate on uncovering behavioral trajectories, prototypical functions, tasks, and action sequences taken by performers; as indicated earlier, whereas CE attempts to uncover the underlying characteristics of (problem solving) work in the domain of practice, including the factors that establish critical boundary conditions on work. CE deliberately avoids the temptation to reduce work to fixed behavior patterns based on the observation that experts adaptively adjust behavior at the time of work to achieve mission goals. As a result, the CE approach attempts to provide a useful workspace that makes the dynamics of the problem visible, at multiple levels of abstraction, and that naturally supports adaptive work. To achieve these design goals a different slant is taken on mission/work analysis.

4.5 Approach to decomposition. The last contrast deals with the issue of decomposition. Decomposition is a powerful technique that aids both problem understanding and design development. A mission model is decomposed in an effort to identify requirements that serve to stimulate the development of a system model that, in turn, is then decomposed into smaller units that can be analyzed and developed in parallel by specialized design teams. This process of breaking down a complex problem into its constituent parts is a fundamental technique in all large-scale system developments- i.e., missions are decomposed into tasks and information requirements, and system work is decomposed into CVI work and aiding subsystem work. Both of these aspects of decomposition are elements of current practice.

The importance of decomposition is not diminished when CE design practices are employed. However, there is an important difference in the processes of decomposition. The differences can be very subtle and thus can easily be over looked. These differences revolve around the type of relations that are preserved in a decomposition. Implicitly, CE has a strong view on this issue. When systems are decomposed into subsystems and problems are decomposed into task sequences, these are examples of decomposition that follow a whole-part relation. Some larger item or system is parsed into a set of smaller ones. Thus, a system is reduced to a set of subsystems that , in turn, are reduced to assemblies, circuits, components, etc.. In like manner, a user interface is reduced to a layout containing displays, formats, symbols, etc. This one (aggregation) dimensional approach to system decomposition provides a design analysis and solution path for large scale development efforts, and helps the system manager to plan, structure, and distribute design and development work.

There is no question about the value of whole-part decomposition. This form of decomposition, however, has the tendency to gloss over important interactions and emergent property aspects of a system. Prevailing design practice addresses this issue mainly through a series of progressive build cycles that include an integration step near the end of each cycle. Problems in integration reveal important interactions. It is also near this integration time that the CVI team has enough information about the capability of each subsystem that it is able to finish the CVI design. In other words, the "human factoring" of the CVI logically follows

the rest of the development. To avoid this situation and for other reasons, the CE approach uses a two-dimensional framework for decomposition from the beginning of problem (understanding) analysis. The additional dimension preserves a nesting relation while decomposition proceeds along an abstraction axis. The abstraction axis or dimension is characterized by a progressive shift in descriptive language from purely symbolic terms to one that uses terms that map directly to physical, concrete items. It preserves a nesting relation, as opposed to the whole-part relation preserved by the aggregation dimension.

Another important factor is to recognize that the focus of the abstraction dimension is on the work problem to be solved, whereas the aggregation dimension focuses on the system to be used to support the user in solving the work problem Together the two dimensions are used to define a conceptual workspace for the system user(s). A diagram of this workspace framework is shown in Figure 4.

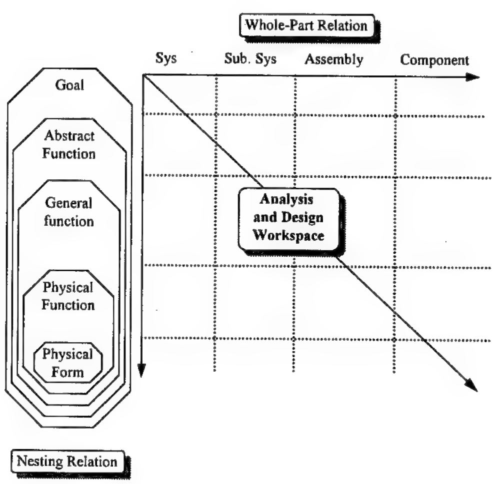


Figure 4. Design workspace defined by an abstraction hierarchy (nesting relation) and a system aggregation dimension (whole-part relation).

Decomposition in abstraction preserves a nesting relation. Nesting refers to the relations between alternative descriptions or description languages of a whole entity or system. It implies that one whole system description is in some sense contained in another whole description of the same system. This is the type of relation expressed in the Abstraction Hierarchy advanced by Rasmussen (1983; see also, Rasmussen, 1986, Rasmussen et al., 1994). The highest level of the framework describes the work problem in terms of general, abstract goals to be achieved by work (see Figure 5). The goal of work may be to win the war, a battle, or a mission, or it may focus on support functions like the logistic function of providing the right material at the right time in the right place.

At the next lower function level, called the Abstract Function level, an attempt is made to express the goal in terms of fundamental principles of the work problem for a particular domain. For example, given the goal of winning a war, the Abstract Function description would be based on some accepted principles of warfare. US Air Force doctrine (Air Force Manual 1-1, 1992) identifies nine such principles. Together, they produce a first-principles model of work. Thus, the language at this level provides a highly conceptual view of the work problem. It sets the foundation or framework for the discovery of relevant work semantics (i.e. description of the work problem) at progressively more concrete levels of description. In some instances the representation at this level can be formed into a dynamical model that provides a compact synthesis meaningful to an expect. Other times the network of constructs captures the performance trade-off space and thus also serve to reveal useful measures of work performance.

ABSTRACTION HIERARCHY

PURPOSE: Categories refer to properties of the environment

ABSTRACT FUNCTION: Categories refer to theoretical constructs that capture problem dynamics

GENERAL FUNCTION: Categories refer to system functions and coordinatio processes

PHYSICAL FUNCTION: Categories expressed in terms of physical processes

PHYSICAL FORM: Categories expressed in terms of objects, appearances, a locations

* Adopted from Rasmussen et al, 1994

Figure 5. The Abstraction Hierarchy.

The next lower level is called the General Function level. It re-expresses the work problem in terms of work processes that, in toto, are expected to successfully solve the work problem. This level supports an input-output functional model description that indicates the coordination needed to operate on the dynamic problem. The functional break out is motivated from an understanding of the fundamental nature of the problem derived from the model developed at the Abstract Function level, which incidentally justifies the level 2 name. From this point on, the subsequent descriptions capture the problem in system terms, with the expressions using

progressively more concrete objects. The forth level, called the Physical Function level, presents a functional model of the physical processes of the system. And the lowest level, Physical Form, identifies the objects by name, and may include their actual appearances, location, etc. [see Flach et al., 1998 for an example of different levels of description and a discussion of the Abstraction Hierarchy in the context of a SEAD (suppression of enemy air defense) mission].

It should be clear that through the use of different types of constructs to model the entire system the Abstraction Hierarchy (AH) preserves a nesting relation while simultaneously "decomposing" the design problem. Decomposition under this relation, of course, is different that under a whole-part relation. Here decomposition is in the sense that as one moves from an abstract problem description, it is decomposed into a more concrete physical one. It is also a decomposition in that the language moves from environmental terms (problem in the world) to theoretical terms (basic conceptual problem) to system process terms (design problem). The whole problem is preserved under the nesting relation; whereas it is parsed under the whole-part relation. When the two forms of decomposition are combined, it is easy to see that the resultant design space helps the designer to view aspects of the system (i.e. focus in a region of the whole-part dimension) relative to the whole system and work problem (i.e. focus at a level of the AH). The contribution of this work space is that it keeps visible (often in invisible ways!) three inter-related views of the design problem (the domain problem, the complete system, and the immediate, narrower design issue). Through the use of this model space the designer is better able to envision design solutions that successfully address the interactions that result from all of these concerns. In this way the CVI designer can use CE as a means of crafting an interface that defines a problem-focused workspace rather than one that aims to support specific work trajectories.

It is interesting to note that Rasmussen (1986) has found that when trouble shooting a diagnostic problem experts tend to move along the left-right diagonal of the workspace. There is some evidence to suggest this is a general finding, at least for diagnostic tasks. It could easily pertain to the work of combat pilots, as well. Pilots must also frequently shift conceptual framework and focus to deal with co-occurring events. If the diagonal captures expert behavior, then the crew vehicle designer may be well served to concentrate on this region of the problem space when making design decisions.

One last comment is in order on the Abstraction Hierarchy. As indicated earlier, the nesting relation allows the problem solving focus to be persevered as functional details of the system are added. The current practice starts with a focus on the problem to be solved but with decomposition the focus shifts progressively to an information processing view, which is further encouraged by activities (e.g. workload analysis) that invite analysis of tasks and information in relation to cognitive demand. In effect, these analyses reduce problem solving to information processing. CE avoids this reductionism through nesting in the functional abstraction decomposition. Information processing is not ignored. Rather, it is addressed as part of an integrated analysis that maps internalized information processing to externalized problem solving (e.g. see the Decision Ladder and the skill, rules, knowledge work strategies described by Rasmussen (1986, 1994).

The characterization of CE presented here certainly is not complete. I have said nothing, for example, about the details of design synthesis for interface design, nor have I addressed evaluation methods. In fact one of the significant strengths of the CE approach is that it does a good job of bring these three aspects of design together in a single framework. (The interested reader should review Rasmussen et al, 1994). Hopefully, the presentation has been sufficient to make clear there are conceptual and practical differences between current CVI and CE practices, and thus has provided at least a bit of insight into the basis from which they have been derived.

5. OTHER CONTRIBUTORS TO CONFUSIONS ABOUT CE

Before we turn our attention to a brief review of weakness in CVI current practices and application issues for CE, it may be helpful to address some other factors that tend to increase the difficulty for system managers and design engineers to understand what is new and different with CE. The origin of the problem stems from the use of new terms that may not be well defined, and from employing new meaning for old terms. This problem occurs because CE is an emerging area of research and, like any other active field, there is a range of viewpoints. In short, there is an intellectual debate about

what CE is, what it should be, and how it should be practiced, and this, in turn, results in a proliferation of new terms and nuances in meaning. This can create a minefield for the uninitiated.

One version of this problem occurs when researchers in the field highlight factors as if they are original concepts that follow from the new perspective of the new field. Sometimes these new insights have already been made independently, and perhaps in a different form, by current design practitioners, and thus have been adopted as part of the prevailing practices. For example, sometimes presentations on CE point out that the use of scenarios is valuable for guiding design decision making, and they advance the concept of scenario-based design. The crew station designer reads this and agrees, but asks: So what is new? Scenarios have been a valuable aspect of crew station design for a long time. The real point may be not so much that scenarios are used or not used but, as we have seen, CE looks at scenarios to extract other forms of information that generally is relatively neglected by prevailing practices. Thus, CE looks to analyze scenarios for factors that help them to understand expert behavior and factors that contribute to error induction. CE attempts to construct scenarios that bring these factors out into the open. This could be a real departure from current practice, but it may be missed by how the concept of scenario-based design is presented.

Another example deserves mention. The concept of Cognitive Task Analysis (CTA) is often associated with CE. It is sometimes advertised as a new and perhaps better way to analyze work. But here too, confusions abound. Sometimes CTA is treated as if it is (nearly) synonymous with CE, sometimes it is cast as a CE analysis method, an analysis framework, or an approach to collect data from the field. Further, some use the term Work Domain Analysis to mean the same thing as Cognitive Task Analysis, but other times they are given similar but slightly different meanings. To complicate matters more, both terms may be used by researchers who hold different orientations toward the concept of cognition. As indicated earlier, there is in-the-head cognition and in-the-environment cognition.

Crew station designers have recognized for a long time the need to improve the inclusion of cognitive work in system analysis as the extent of automation has been growing in military systems. They have long used domain experts in the design process to help elucidate the work domain, to pin point difficult work problems, and to suggest and critique design concepts. They have performed various types of cognitive analyses. It may be difficult, therefore, to see how CTA differs from current practices. The differences may be in subtle ways, like what the analyst probes for or how the information is used once collected. In CE, Cognitive Task Analysis feeds the construction of the Abstraction Hierarchy. This is different from current practice, and it may have an important effect on design. Unfortunately, it may take more effort to discover these differences than system managers and practitioners are willing to invest to determine if the new research has anything of value to offer.

6. WEAKNESS IN THE CVI DESIGN PROCESS

Large scale system design is a complex sciotechnical process. The methods for carrying it out have evolved over many years. The simple fact that the military has been able to turn out sophisticated human-machine systems that work serves to validate basic elements of the process. But this is not to say that it is perfect or static. For example, there have been notable difficulties encountered with the design of software-intensive systems. There have been difficulties with systems that certain a significant level of aiding automation. There is always room to improve design process. This is true in the crew station area as it is in others. Here I will mention what I regard as four pervasive weaknesses of current practice.

Mission models are expanded as needed to support specific design decision making in each subsystem area. Two potential problems can arise from this. First, details of the models used at design time may be different from those used in analysis, even though all models may be indexed to the approved composite model established in the early stages of a project. Second and relatedly, each subsystem design team may expand a mission model as needed to guide local design decision making. All of these locally different models tend not to get coordinated and reflected back into the mission model used for evaluation. One result of this is that opportunities and constraints important to CVI design may not be made visible to the CVI design team. This can occur with the CVI team itself, as well as with its interactions with other subsystem teams. This can also contribute to integration problems.

A second area of concern addresses a problem of circularity that is inherent in the process of determining tasks and information requirements. An attempt is made to keep mission and task scenarios "technology free." The goal is to define the requirement functionally, and then appeal to technology as the solution. Unfortunately, this ideal cannot be achieved. Implicit assumptions that are technology based always sneak into the "technology free" analysis tools. It is not possible to provide a meaningful (time-index) task description that does not contain assumptions about technology used in the task. The spatiotemporal properties of a task are derived from the problem situation *and* the resources available for an actor to work on it. Both the nature of the work and the time it takes to apply a resource is, implicitly if not explicitly, tied to assumptions about the technology. This, of course, makes the requirements generation/functional description process circular. This circularity can be broken, up to a point, by using concepts at different levels of aggregation or abstraction, but this introduces other problems.

The quest for the development of a robust CVI is another area of concern. It is well recognized that the mission conditions under which military systems will be used are in a constant state of change. It is also recognized that a system must be used flexibly and adaptively by the human user. Indeed, human adaptivity is critical to the success of the system. A CVI that constrains adaptive performance reduces the utility of a system. Thus, the goals is to build robust CVIs, ones that can be used easily and effectively to meet the exigencies of the current situation. Current design practice has trouble in this area because it uses fixed behavioral trajectories as the basis for CVI design. Flexible solutions can be achieved when design is approached in this manner by using principles of multiple input paths and common input methods, but these solutions often contribute to interface complexity and may at times hinder adaptive performance.

Another weakness in current practice relates to incremental design. This also extends beyond the CVI area. On large scale system development efforts, incremental design often occurs whether or not a prototyping model is used. Incremental design results in functional detail being added progressively. Generally, basic functionality is sketched out, and details are inserted over successive build cycles. Usually the list of functions to be included in the system are approached more or less from the easiest to the most difficult, with difficult items held off until late in the project. This provides more time to work on hard problems, but if easy problems turn hard, which always seems to happen, then some of the hard problems fall off the design bench. An alternative is to design by behavioral threads, where each thread covers a work problem. Note this is not the same as designing behavioral trajectories. Threads contain all factors necessary to reveal the dynamics of a work problem, whereas design by trajectory provides a single paths solution that often does not generalize well to variations in task conditions; hence it is brittle.

Finally, I reiterate the fact that CVI activities are not well integrated with the design of user-aiding subsystems or even new interface technologies. While these technologies require a user-interface to show their functional contribution, the CVI design team is only asked to "human factor" the interface for these systems after a preliminary interface is provided. Thus, the CVI designer rarely knows about the detailed dynamics of the aid. This knowledge is needed to construct a good CVI. Therefore, current practice indicates at least a partial disconnect between mission, task, cognitive, and workload analysis activities and CVI design activities, since de facto aspects of CVI design are handled in unknown ways by other groups, including device vendors.

CE has not emerge as a way to specifically address these issues, but it should be easy to see from the previous discussion of CE practices that it may provide a means to mitigate at least some of these problems.

7. APPLICABILITY OF CE TO MILTARY SYSTEMS DESIGN

We have seen that CE has some distinguishing characteristics that are generally not included in crew vehicle interface design for military systems. These represent advancements made by the CE research community, and they have been applied to the design of control rooms for power plants, medical devices, and stand-alone decision support systems. The question is whether or not they can be applied with equal success to fully integrated military systems. Some hurdles may have to be overcome. First is the question of scale to meet the challenges of military problems. Previous successes have been in domains where the focus of (work) problem solving activities has been on understanding what the system is doing and then making adjustments to it, or simply using the device

itself. Thus, in some sense, the "mission" problem and use of the system are isomorphic. This degree of isomorphism does not hold for military systems. The mission problem is external to the to-be-designed system and it can change radically even though the system does not change. How well can CE handle this type of situation?

A second aspect of the scale up problem deals with adaptive aiding. Adaptive aiding means that the support system itself changes while the actor is solving the external problem. This introduces more change and sets up the need for a dialogue between the actor and machine agent(s). CE has not had to deal with this problem in previous design work, but the problem has been noted in the commercial aviation domain.

Finally, the sheer number of variables that need to be considered would seem to be greater than anything that has been tackled to date. Can CE scale up?

8. COGNITIVE ENGINEERING: FAD OR STEP FORWARD?

CE is an emerging set of design engineering practices and a theory of design aimed at producing more robust user support systems without the need for costly redsign. The field is still in a early stage of development. Naturally, those in the field have differences of opinion about aspects of theory and emerging practices in terms of their ability to support and guide design synthesis. Even though there are differences, I believe all CE theoreticians and advocates would accept the five factors presented here as characteristics of CE. Like me, they would probable consider these factors as a growth in human-centered design science that builds on pervious advances derived from human factors, cognitive science, industrial psychology, industrial and systems engineering, and other disciplines. The distinctions that have been made here between the use of CE in design and current crew station design practices is an attempt to show that there are important differences, albeit in many cases subtle ones, that could contribute to improvements in the design of crew vehicle interfaces as work support systems. Is CE a passing fad or is it real progress? The answer to this question depends at least in part on whether or not CE researchers can demonstrate how CE makes valuable contributions to design, and how it can be integrated into the larger system design process. Crew system designers, aiding subsystem designers, and system development managers must all perceive enough value in engaging in CE practices before it will be adopted. Clearly, CE advocates can help their case by producing design artifacts that other subsystem designs value and want to use. Fad or not? Only hard work and time will tell.

9. REFERENCES

Aldrich, T.B., Szabo, S.M., and Beirbaum, C.R. (1988). The development and application of models to predict operator workload during system design, in G. McMillian, D. Beevis, E. Salas, M. Strub, R. Sutton, and L. Van Breda (eds) *Application of Human Performancce Models to Systems Design.* New York: Plenum Press.

Flach, J.M., Eggleston, R.G., Kuperman, G.G., and Dominguez, C.O. (1998, in press). SEAD and the UCAV: A Preliminary Cognitive Systems Analysis. *Air Force research Laboratory Technical Report,* WP-TR-1998-0013.

Flach, J.M. and Dominguez, C.O. (1995). Use-Centered design: integrating the use, instrument, and goal. *Ergonomics and Design,* pp19-24.

Hollnagel, E. and Woods, D.D. (1983), Cognitive systems engineering: new wine in new bottles, *International Journal of Man-Machine Studies,* 18, 583-600.

Hutchins, E. (1995) *Cognition in the wild.* Cambridge, MA: MIT Press.

McCracken, J.H. and Aldrich, T.B. (1984). Analyses of selected LHX mission functions: Implications for operator workload and system automation goals. Technical Note ASI479-024-84. Army Research Institute Aviation Research and Development Activity.

Norman, D. (1984) Cognitive engineering principles in the design of human-computer interfaces, In G. Salvendy (ed.), *Human-Computer Interaction,* Amsterdam: Elsevier Science Publishers, 11-16.

Norman, D. (1987) Cognitive engineering--cognitive science, in J. Carroll (ed.), *Interfacing Thoughts: Cognitive Aspects of Human-Computer Interaction,* Cambridge MA: MIT Press, 325-336.

North, R.A. and Riley, V. A. (1988). W/INDEX: A predictive model of operator workload, in G. McMillian, D. Beevis, E. Salas, M. Strub, R. Sutton, and L. Van Breda (eds) *Application of Human Performancce Models to Systems Design.* New York: Plenum Press.

Roth, E. M. and Mumaw, R.J. (1992) Cognitive engineering, in (J. Marciniak (ed), *Encyclopedia of Software Engineering,* Vol 1, 110-124.

Roth, E.M. and Woods, D.D. (1989), Cognitive task analysis: an approach to knowldge acquisition of intellgent system design, in G Guida and C. Tasso (eds), *Topics in Expert System Design,* North-Holland, New York, 233-264.

Rasmussen, J. (1983). Skills, rules, and knowldge; signals,signs, and symbols, and other distinctions in human performance models, *IEEE Transactions on Systems, Man, and Cybernetics,* 13(3), 257-266.

Rasmussen, J. (1986). *Information Processing and Human-machine Interaction: An Approach to Cognitive Engineering.* New York: North-Holland.

Rasmussen,, J., Pejtersen, A.M., and Goodstein, L.P. (1994). *Cognitive Systems engineering.* New York: John Wiley & Sons.

Vicente, K.J. and Rasmussen, J. (1992). Ecological interface design: Theoretical foundations, *IEEE Transactions on Systems, Man, and Cybernetics,* 22(4), 589-606.

Woods, D.D. and Roth, E.M. (1988). Cognitive systems engineering, in M. Helander (ed.) *Handbook of Human-Computer Interaction,* Elsevier Science Publishers, North-Holland.

Woods, D.D. and Roth, E.M. (1988). Cognitive engineering: Human problem solving with tools, *Human Factors* 30, 415-430.

IDENTIFYING THE SOCIAL AND COGNITIVE REQUIREMENTS OF TEAMWORK USING COLLABORATIVE TASK ANALYSIS

M. D. McNeese
United States Air Force Research Laboratory
Collaborative Systems Technology Branch
AFRL / HECI, 2255 H. Street, Bldg 248
Wright-Patterson Air Force Base, OH USA 45433-7022

J. R. Rentsch
Wright State University
Dayton, OH

SUMMARY

Typically, collaborative systems design only considers the technological imperative while ignoring the social, cognitive, and contextual components of teamwork. This paper describes a basis for addressing the social and cognitive requirements of design, given a situated problem context, through the use of collaborative task analysis. Inherent in the approach taken is the goal of eliciting, assessing, and measuring *team schema* that influences team performance in complex operational systems. Several real world and operational examples are described to highlight these aspects of teamwork.

INTRODUCTION

Clearly, the number of crews, teams, multi-operator units, and collaborative systems within the Air Force and other military organizations is expanding significantly. As missions and tasks increasingly become reliant on joint problem solving and teamwork, there is a growing necessity to identify social, cognitive, contextual, and technological requirements. Increasingly, emergent complexity underlies many of the difficult problems that demand teamwork. Missions emerge in a way that makes them ill-defined, uncertain, risky, and full of tradeoffs with differing consequences. Team members strain to make sense of these situations, reacting to external stimuli and interacting with each other to interpret their environment, under stressful conditions.

One approach to facilitating teamwork in ambiguous situations is to introduce technology. For example, video teleconferencing and groupware systems may be used to afford high level decision making / coordination responses when teams are remotely distributed in geographic space. In the *New World Vista* reports (Ref 1), technologies such as avatars, shared virtual environments, intelligent social agents, and mobile ubiquitous computing are all promised as forms of automated assistance and decision support to enhance teamwork. Yet, if these complex systems and information technologies are designed without being informed by an understanding of the cognitive, contextual, social, organizational, and cultural components that predicate work, failure is certain to ensue.

In this paper we suggest a proactive approach to emergent complexity and teamwork by examining various techniques to elicit the collaborative basis (e.g., knowledge, skills, mental models, procedures, domain practice, and constraints) of group sensemaking (Ref 2). As Jens Rasmussen has eloquently pointed out interdisciplinary problems demand solutions that consider the vertical integration of constraints (e.g., socio-political, organizational, regulatory, team, and individual variations). The value of these techniques may be assessed in contrast to traditional individual-level task analysis which has been used in (1) human factors to assist in the design of human-systems interfaces and (2) industrial / organizational psychology to evaluate team jobs. The paper describes methods that the authors have created and used that are predicated on the concept of *team schema*. Finally, the paper discusses where there is room for improvement and what is yet needed in this area of endeavor. First, several examples of real world problems are given to anchor the overall discussion.

Situated Problems

The nature and type of problems experienced by teams in operational settings are naturalistic and involve multi-faceted elements that emerge in situations. Suchman (Ref 3) has referred to these as situated problems. Young and McNeese (Ref 4) explain situated problem solving from an ecological perspective. Many everyday experiences can bring forth situated problems such as the two presented below.

In-flight emergency. I recently experienced a situated problem while flying home from a NATO meeting. The jet I was on was over the middle of the Atlantic ocean when one of the passengers stopped breathing and passed out. Immediately, the routine operations of the flight crew and passengers took on a new social, cognitive, and emotional aura. The time pressure to act immediately was intense. Everyone in the vicinity of the passenger was under differing forms of stress. I was about three seats away from the ailing passenger and I feared that she might loose her life in flight.

Because the plane was over the ocean (thereby landing near a hospital was not possible) situated problem solving involving collaborative activities had to quickly develop within the constraints of the resources available within the flight. At first the stress was paralyzing but then the gravity of the situation was realized and people worked together as a newly formed team consisting of the flight-crew members and passengers. This situation -luckily- was assuaged as a doctor and two nurses were on board. The person had complained to the flight attendant that she was dizzy and loosing consciousness. By the time the attendants, co-pilot, and medical personnel arrived at the passenger's seat she was out cold. At this point, the nurse could not find a pulse. With the help of adjacent passengers, the afflicted person was pulled out of her seat-belt, lifted out of her seat, carried to the back seats of the plane, and laid out across several seats. By this time, the doctor had opened his medical bag and attended to the patient.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

Thankfully, he was able to help resuscitate her and to get a pulse. Simultaneously with the doctor working with the patient, the rest of the newly formed team readied the oxygen apparatus for administration to the passenger. The remainder of the flight (approximately three hours), the passenger stayed on oxygen and continued to lay down (and sleep) in the rear of the plane. I assume she was taken to a hospital upon arrival at our destination.

Perhaps this problem is extreme but it shows some of the complexities, interdependence, and constraining conditions that must be considered simultaneously - often in short order - in situated problems. More examples of naturalistic problem solving (and approaches that can be considered) can be found in Zsambok & Klein (Ref 5). Beyond the situated problems that exist in everyday activities are problems that evolve within complex operational systems. Take for example the Mir Space Station.

Mir Space Station. There have been a number of incidents and emergencies that have evolved over the last year within the Mir space station, a very complex operational system that requires joint problem solving. The crew has experienced emergencies such as a fire, leaking oxygen, a collision, significant loss of power, complete darkness, and free drift. These examples that show the interconnected nature of problems, how problems lead to unanticipated subproblems, and the consequent need to focus on multiple cognitive processes to effectively problem solve. In addition to requiring the cognitive components inherent in problem solving, the Mir missions have stressed many of the social, cultural, and unusual aspects of joint problem solving. For example, there are cultural differences in risk taking between the Russians and Americans. Most recently, an astronaut and cosmonaut engaged in a collaborative space-walk plagued with challenges of communication (i.e., the language difficulties had to be assuaged for joint problem solving to ensue). Add to this the exhaustion and stress induced by space travel and the health problems (heart condition) experienced by the Mir commander, and we see the many ways a complex system can go awry.

An Integrative Construct - Team Schema

Given the emergent complexity experienced within situated problems (such as the ones just elaborated), a conceptual construct that integrates multiple cognitive processes, individual-to-team performance, and social factors is *team schema*. This may be roughly calibrated to Rasmussen's concepts of skills, rules, knowledge; signals, signs, and symbologies (Ref 6) and their use as potential models of the human in work systems. The concept represents an expansion of the idea of individual schema (or mental models), as a basis to define what a person knows and can know, to that of what the team knows and how effectively it can adapt its actions to contextual boundaries and variation. Kant originally described the schema as a mediating representation that intervenes between an individual's intellectual and sensory experience that actively provides an interpretation of that experience (Ref 7). Bartlett (Ref 8) further made use of the schema as an abstract cognitive structure that is activated upon interaction with the environment and is representative of prior information that is organized in specific ways. These early schema explorations provide a historical basis for development of the team schema construct.

We will now look at how the team schema construct can be applied to understand how individuals work together to fuse their individual skill and knowledge to solve situated operational problems.

Complex decision making and joint problem solving in Battle Management Command, Control, Communication, and Intelligence (BMC^3I), requires timely integration of team member's cognitive processes, knowledge, skills, and work environments in order that effective teamwork transpires. In some BMC^3I situations (e.g., plan execution), operations require a high level of interdependence, and hence group interaction, whereas other BMC^3I work elements are loosely coupled (e.g., information analysis) in that they benefit from joint collaborative activity but are not necessarily predicated upon collaboration.[1] Underlying successful teamwork integration in this context is a person's teamwork schema.

The teamwork schema is a cognitive structure that organizes an individual's thoughts about how teamwork transpires (a kind of team mental model) and it acts as a basis for assimilating and making sense of teamwork situations. Teamwork schemas, which are schemas that contain information regarding the processes by which team members interact, communicate, and complete the team's work (Ref 9), have received some research attention (e.g., Ref 10, 11). Teamwork schemas contain knowledge and information relevant to communicating about, evaluating, and compensating for teammates' performance. Teamwork schemas will guide team members' assumptions, expectations, and behavior regarding the process of working together as a team.

Team Member Schema Similarity (TMSS) refers to the degree to which team members' schemas overlap (Ref 12). In particular, TMSS may refer to schema accuracy as well as schema agreement (Ref 10). Schema accuracy refers to how accurately team members can describe one another's schemas. Schema agreement refers to the degree to which team members schemas are similar.

Team effectiveness is a critical factor in many Air Force problems such as unmanned air vehicle operations, information warfare, BMC^3I, and joint space systems. It is hypothesized that optimal levels of team member schema similarity will predict team effectiveness. Team members who have schema similarity have similar knowledge about teamwork and they organize this information similarly. Teamwork schema similarity is hypothesized to enhance team effectiveness because similar teamwork schemas in terms of schema agreement and schema accuracy among team members will allow team members to interact efficiently and effectively. Many difficulties may occur when team members' teamwork schemas are not similar. In contrast, when team member schema similarity is high the team shares knowledge effectively therein leading to a high level of team performance. Team members with high TMSS will be able to anticipate, facilitate, and compensate for one another's behavior. In addition, similar teamwork schemas may reduce process losses typically associated with teams (Ref 13). Communication among team members is also likely to be enhanced as team members' teamwork schemas become increasingly similar. Team members may be aware of the

[1] Personal communication with Dr. Robert Eggleston, Technical Director, Collaborative Systems Technology Branch, U. S. Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio.

information required by each other and fully understand the information that is being communicated to each other. Moreover, team members are likely to anticipate and understand each other's actions when TMSS is high.

With the advent of many advanced information technologies in BMC^3I settings, the understanding and measurement of team member schema similarity is expected to be a prime variable to predicate effective collaborative systems design. This is particularly the case for geographically distributed teams.

METHODS OF COLLABORATIVE TASK ANALYSIS (COLLATA)

Elicitation and Assessment of Team Schema

There are several methods that have been utilized to elicit the social and cognitive requirements associated with teamwork in situated problem domains. Some of these methods emanate from the human factors / cognitive engineering perspectives whereas others are more akin to the job analysis area of industrial / organizational psychology. Vicente (Ref 14) provides a historical exegesis of task analysis, cognitive task analysis, and cognitive work analysis while describing the differences among these methods. All of these methods tend to echo trends in work itself with early task analytic approaches reflecting manual physical labor practices. More current cognitive analyses tend to relate to the need to address user's conceptual knowledge and cognitive skills as they interact with systems. A key point that Vicente makes is that many of these methods *only* consider tasks that are fairly well-defined and anticipated. The take-home point is that situated work problems are often ill-defined, unanticipated, contextually-bound, and involve collaborative-social activities; in addition to requiring cognitive considerations. For example, the in-flight emergency presented earlier in the paper, is like this. Hence, Vicente suggests that looking at *tasks* is too limited and that a broader approach that considers work domains is needed (i.e., cognitive work analysis). The Rasmussen framework (Ref 15, 16) is an example of cognitive work analysis. Within this framework social and organizational *constraints* influence and shape work. Subsequent use of this framework has often focused on organizational shaping constraints as they relate to the individual worker - rather than collaborative teamwork activities. This paper emerges from the Rasmussen perspective but places more emphasis on the *social construction of knowledge* as it influences: a) collaborative activities in situated problems and b) the consequent development of team schemas; to propose a new collaborative task analysis[2] perspective.

Situated problems require joint articulation and coordination of team members' schemas in order to effectively share teamwork tasks, performance, decisions, and constraints. Theoretically this aligns with the social construction of knowledge and shared cognition literature (Ref 17) that reify the view that knowledge is an interpretation of social experience, and that these interpretations also enable and constrain individual processes of sense-making.

Social constructions of knowledge can be obtained in accordance with the comprehensive cognitive engineering process outlined by McNeese (Ref 18) and may ensue by (1) examining protocol transcripts (2) actively eliciting team schema (3) assessing decision ladder / abstraction hierarchy representations (Ref 16) (4) measuring team member schemas similarity (Ref 11) in joint cognitive systems to see how collaborative performance differentiates across entity, goal, form, function, function allocation, and action (5) utilizing cognitive engineering data analysis techniques (Ref 19) for observed behaviors. Although all these areas of CollaTA have been under study in our laboratories, the two described here are group concept mapping and team member schema similarity measurement. Concept mapping is a highly participatory-conceptual technique that focuses on eliciting schemas that a team uses to solve problems, whereas the TMSS focuses on empirical-experimental measures of team members' schemas.

Group concept mapping. McNeese, Zaff, Brown, Citera, & Wellens (Ref 20) indicate that the issues that often create problems or predicaments for effective teamwork are also present when eliciting knowledge or analyzing work tasks in a group setting (i.e., group concept mapping). They suggest four operative difficulties that can incur: 1) the group may have access to different information which has not been distributed equally across the group, 2) the group members may have access to the same information but interpret differently, 3) the group may lack a common conceptual framework (in other words their team schemas may be under-developed), and 4) team members may fail to communicate their perspectives to other team members. In order to assuage these conditions, they have recommended a user-centered, participatory approach to design effective decision support / collaborative systems technology; wherein external representations of the schema underlying performance are elicited. One method they have utilized, concept mapping, is a process by which an individual's underlying knowledge (expertise) about a task, procedure, context, work, or teamwork is elicited by a mapper (or a team of mappers). Expertise is represented as an interconnected network of concepts and relationships between concepts as it elicited first from an *individual* subject matter expert. A 'concept map' affords (for an expert and a mapper) a shared communication medium to develop in-depth relationships about the complexities (and potential problems/difficulties) of a specific aspect of their work. The approach requires individuals, who are intimately involved in a given collaborative work domain, to describe the constraints of their work from their own perspective which would include ways by which teamwork could be made easier, more efficient, and more effective. Concept mapping (at this individual level), is hence indicative of a cognitive task analysis and primarily elicits declarative or intuitive knowledge through a picto-literal schematic representation. Various cases have been studied using concept mapping to distill an expert's model of a situation (Ref 21).

However, to address the nature of teamwork and the inherent difficulties and issues, McNeese and his colleagues (Ref 21) have explored various dimensions of Group Concept Mapping (GCM) as a unique form of CollaTA. GCM requires both individual and group perspectives of collaborative work to ensue. The individual procedure as indicated above assesses each team member's understanding of their own schema and the team schema as they work together with other members. This first level procedure is done separately and independently from the group (as a

[2] The authors considered changing this name to collaborative work analysis to correspond with Vincente's (1995) arguments but decided to keep the task moniker to maintain the historical connectivity with task analysis.

whole) and is completed for each member on a defined team without influences from other members. However, it is also necessary to distill a more global picture of group activities. Because each individual's view of teamwork is likely to be prejudiced by his/her own personal experiences, a group procedure is also part of the GCM process. This takes the form of a collective interview (while still utilizing the concept mapping elicitation and representation props). As a GCM is developed with all the members present and simultaneous participating in building the map, we have observed a process of team situation awareness occur where members become aware of other member's requirements, tasks, constraints, preferences, biases, roles, and interdependencies. In some cases, these social and cognitive requirements were known previously but in many cases the group process elicited new information that was enlightening for other fellow team members. The GCM creates an atmosphere where collective induction can occur among members thus resulting in the elicitation and assessment of team schema. The group procedure also reveals insights about how teamwork can be done more efficiently and effectively among the members, hence highlighting the discover aspects of collective induction.

In summary, we have found that our use of GCM provides a shared communication medium that facilitates explicit representation of team member's differing perspectives on situated problems. It acts a perceptual anchor and group external memory aid for participants. A GCM allows for members to follow a string of ideas and then return to original points for further elaboration, assessment, and clarification. It provides a means of visualizing knowledge which enhances the building and understanding of team schema which results in identifying and communication of complex ideas and relationships (Ref 20). To expand GCM as a cognitive engineering technique it has been used in participatory design practice with group design storyboarding to translate team schema information into actual group interface designs. Additionally, much of our future work in this area is based on advanced form of GCM, fuzzy cognitive maps, which are utilized to build dynamic models to simulate team processes and constraints (Ref 22).

Team member schema similarity measurement. The Team Member Schema Similarity Model (Ref 10) is shown in Figure 1. The critical variable in the model is team member schema similarity. Organizational researchers have obtained indirect evidence for a relationship between TMSS and team effectiveness.

Most of the research provides evidence relevant to TMSS conceptualized as schema agreement. The accuracy component of TMSS has not been explored extensively to the authors' knowledge. Studies at the U.S. Air Force Research Lab / Collaborative Systems Technology Branch are presently testing the roles of team member schema accuracy and agreement in the prediction of team effectiveness. Rentsch, Pape, and Brickman (Ref 23) examined TMSS conceptualized as schema agreement and schema accuracy. Their results revealed that agreement and accuracy predicted team effectiveness significantly and that schema accuracy may be a more significant predictor of team effectiveness than schema agreement.

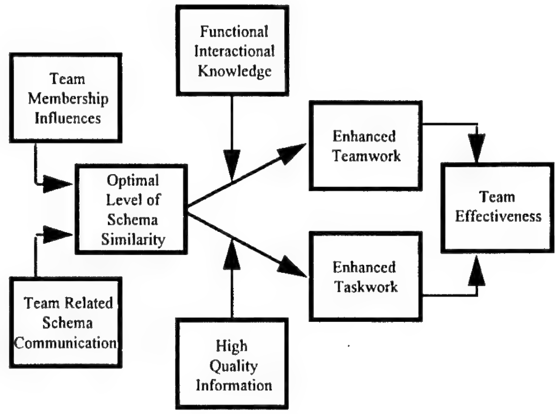


Figure 1. Team Member Schema Similarity Model

As shown in Figure 1, two antecedents of team-related schema similarity are team membership influences, such as person-environment fit, and schema content communication, such as that which occurs during socialization processes (e.g., as occurring in the social construction of knowledge). These antecedents are expected to regulate the degree of schema similarity among team members to an optimal level. It is assumed that an optimal level of schema similarity exists for any given team and that when schema content is of high quality, the optimal level of schema similarity will enhance team effectiveness maximally.

Teamwork TMSS has been measured (Ref 24) using a schema elicitation group interview technique. These data are then utilized to construct a set of stimuli reflecting teamwork schemas, which is administered to intact teams. Each member responds to the stimuli either using a paired comparison method or a standard rating method. These data are then analyzed using a variety of techniques including multidimensional scaling and pathfinder to determine the degree of team member schema similarity. TMSS measurements have been shown to predict team effectiveness in Air Force quality management teams, project teams, and laboratory teams. Research is currently underway at the Organizational Research Laboratory at Wright State University to further investigate the construct validity of the TMSS measurement technique.

The group elicitation method. In addition to the methods we have utilized in our research other options exist for conducting CollaTA. Another method, the Group Elicitation Method (GEM) (Ref 25) is conceptually similar to group concept mapping in that it employs a participatory design methodology to capture multiple requirements from a group of experts / users. Like group mapping, GEM examines CollaTA from the perspectives of: a) the social construction of knowledge and b) participatory design. In many ways one might consider GEM a group brainstorming / decision support tool for meetings because it is embedded within computer software. However, the method itself can be considered a form of CollaTA because it affords the elicitation of design rationale, multi-specialty knowledge from a group of interacting design / usability experts. The method generates a setting for the 'collaborative analysis' of tasks but can be used to analyze 'collaborative tasks' as well. The formulation of knowledge begins through the elicitation of issues via the use

of checklists (e.g., What is the goal of the engineered system?). From issues, group participants generate viewpoints and expand others viewpoints by using a technique termed *brainwriting*. Viewpoints can then be expanded into more elaborate concepts.

The *Cobweb* conceptual clustering system (Ref 26) is used to reformulate and combine views. Concepts and their inherent relationships with each other are rated for viability and priority by participants. Once these steps are completed, scores can be derived that measure consensus among participants. Finally, an analysis of the concept network completes the process and reifies the consensus process. Unlike, traditional task-analytic techniques in human factors, GEM incrementally builds a domain's ontology through knowledge confrontation and interpretation, given other team members' multiple perspectives. This procedure is thus a means to afford systematic evaluation of collaborative design / user rationale. GEM has been applied in the aeronautics domain.

Team task analysis. Baker and his colleagues (Ref 27) referred to *team task analysis*, which they define as including an analysis of the team's tasks and the team's teamwork requirements. They cite the FAA as now recommending such an analysis for in the design of pilot training. Baker et al. reported that in team task analysis has been rooted in traditional industrial/organizational psychology job analysis methods, such as critical incidents techniques. They criticize these methods as being singular in their approach and as inadequately addressing all aspects of teams. Although, they lament the lack of research in this area, they provide only a list of questions as the next step.

Bowers, Baker, and Salas (Ref 28) attempted to investigate the reliability and validity of several importance measures for a job analysis in which they assessed team related task information. They examined five importance indices for four task dimension ratings (importance to training, task criticality, task frequency, overall task importance). Three jobs were analyzed in the study. Team tasks were developed for three jobs (42, 56, and 56 tasks), which were rated by 46, 33, and 34 job incumbents. The results revealed that all importance indices had relatively low interrater reliability and validity. Basically, their approach proved disappointing, but Bowers et al. suggested that their results indicated a need to develop a new approach for identifying the importance of team tasks. In addition, they suggested that a new set of task dimensions be developed on which team tasks could be rated for importance.

Team design questionnaire, Campion and his colleagues (Ref 29, 30) have developed a questionnaire for assessing team design characteristics. Although the questionnaire was designed for research purposes and was not designed to conduct CollaTA specifically, it has potential to contribute to the topic. Drawing from social psychology, socio-technical theory, industrial engineering, and organizational psychology, Campion identified 19 characteristics related to work team design. Samples of the 19 characteristics include: self-management, participation, interdependent feedback and rewards, potency, and communication and cooperation within the work group. Sets of three items per each characteristic were developed and the items are rated on a 5-point agree-disagree response scale. Team members rate how descriptive each item is with respect to their team. They reported that 15 of the 19 scales had acceptable reliability.

This type of approach offers an expedient method for obtaining collaborative information. It is not subject to the initial development phase as in the TMSS measurement method and the concept mapping. However, expedience may be traded for information regarding the cognitive aspects of teamwork.

EXAMINING TEAM SCHEMAS IN UNMANNED AIR VEHICLE (UAV) DOMAINS: AN EXAMPLE DOMAIN OF INTEREST

The UAV problem domain, as a specific example of BMC^3I environments, will absolutely need to consider the social and cognitive requirements of work in order to evolve an effective collaborative systems design. The domain is one that interweaves cooperative and individual work, at various levels of emergent interdependent activity. Team members having specialties in navigation, electronic warfare sensors, weaponry, and mission planning must communicate and coordinate a variety of activities while exerting joint remote control over an uninhabited flying vehicle capable of performing various reconnaissance or combat missions. Coordination and control is contingent on noticing, extracting, integrating, and reasoning about multiple information sources that will concurrently vie for a team's attention as the mission situation changes. Given the interdependent nature of collaboration, the UAV team must organize decision making and performance by adaptively employing articulation, coordination, scheduling, alignment, meshing, and integration. These qualities typify situations which the UAV crew will face in function and role integration activities. Complexity is at the heart of such work and in order to address the complexity issues a common ground of understanding must ensue in spite of the ensuing sense of disorder (uncommon ground). Complexity especially seeks out situations that are distributive and interdependent in nature, and in turn these situations require a greater degree of articulation (Ref 31), e.g., the communication of team schema as part of work.

In contrast, situations that involve everyday social interactions through face-to-face engagements are managed intuitively resulting in effective and efficient taskwork-teamwork. So perhaps the ground-based UAV crew who work together in close physical proximity naturally and intuitively articulate common meanings in their work. These type of settings can facilitate the development of teams that are successful rather than teams that fail. Successful teams develop teamwork schemas that are highly similar in nature. Teamwork schemas are a major component of UAV work in accomplishing valuable missions. When team members develop similarity in their schemas, then team members should monitor each other, perform activities that support each others' awareness and understanding of the work, take account of each other's past, present, and prospective activities while planning and conducting their own work (Ref 31).

In a collocated setting (such as a command post war-room), highly similar teamwork schemas may be evidenced via gestures, talk, writing to each other; meshing of activity is dynamic and seamless. Often such activities are overseen, managed, coordinated, and accountable through the commander in charge who may also delegate collaborative tasks as needed. When *remotely distributed teamwork* is necessary to manage and control UAVs, the extent to which interdependent activities are coordinated in space and time can determine the degree to which a team succeeds or fails. As

mentioned previously, when settings facilitate face-to-face interactions, coordination is articulated via everyday social and communication modes (e.g., gesturing and pointing at displays). In this setting, TMSS is likely to develop naturally and intuitively. Yet, when distributed work ensues these modes are severely corrupted from the naturalistic realm and forced through more impoverished interfaces. In both distributed and collocated situations, one most likely encounters some form of coordination artifact (e.g., in AF missions one might operate via rules of engagement and standard operating procedures). Much of our research agenda is projected to investigate UAV team schema to formulate different types of decision support / group interfaces. This is where the preceding CollaTA methods would be valuable to use in understanding the team schema in UAV operations. The core of a coordination artifact must readily be viewed as an organizational construct which imprints a coordination protocol into a distinct artifact which stipulates and *mediates* reduction in the complexity of UAV work as it emerges across roles, functions, and requirements. As a software component, the protocol must reflect the mutual changes enacted and conveyed by the joint cognitive system (Ref 31).

In spite of the levels of mutual dependence required for UAV work to successfully be performed, individual taskwork functions are required as well. UAV work will likely encounter inextricably woven work activities in practice. A crew may shift between streams of activities on both individual, cooperative, and even team-to-team work. Hence, a joint cognitive system must be fluid in supporting and meshing together these elements. The respective abstract function levels within these elements must be integrated at the general system level (Ref 31). One idea that has potential for exploring decision support systems / group interfaces that afford these capabilities is the extension of the current Team Avatar research currently conducted by Max Wells and his associates at the University of Washington (Ref 32). This would represent the *next phase of our research* wherein agents - based on the output of collaborative task analyses take the form of avatars. Avatars could help support perceptual anchoring of articulation work in the remote distributed condition as the complexity necessary to articulate and develop teamwork schemas is immensely more complex. Because articulation work demands that interaction cues be embedded in objects in the field of work, ready at hand, ubiquitous, and constantly monitored (Ref 31), avatars created in a common information space (that adapts information temporally and spatially) that can be seen by other 'team members' are projected to have high exploratory value in producing successful teams (owing to an increase in TMSS). This forms an interesting architecture for modeling and enhancing UAV work for future AF missions. It also exploits past work in ubiquitous computing for establishing new formalisms of group interaction (Ref 33). Furthermore, the avatar work establishes the kind of bandwidth and anchoring necessary to enact articulation as *visualized coordination* wherein the coordination mechanism itself can be at once seen while negotiated to assuage incoming change. This would sustain the charge that Bannon (Ref 34) has made that artifacts both shape and are shaped by the participants engaging them; and consequently reflect these changes in individual and team schema.

DISCUSSION

There is much promise in the application of collaborative systems to situated problems such as the use of UAVs in advanced missions. Yet, all is not well. Simply hoping that advanced information technology is the elixir for collaboration problems is misguided. Use-centered collaborative systems design will likely require analyses that delineate and assess the conceptual, empirical, and symbolic elements of team schema, especially when work requires the integration of distributed operations, information technologies, and intelligent agents. This paper has presented group concept mapping and team member schema similarity measures as potential methods of CollaTA, useful for eliciting, assessing, measuring, and evaluating team schema in situated problem domains that involve complex operational systems. The UAV domain has been examined as an example of where consideration of team schema - as an integrated construct - would be valuable. What remains is to (1) continue to expand the use of these methods in various real world domains to analyze the social and cognitive requirements of teamwork - 'collaborative task' analysis while simultaneously exploring (2) mature developments of the participatory collaborative 'task analysis' using meeting support methods such as advocated in the GEM system. Taken together, these directions would broadly define 'CollaTA as the *collaborative analysis of collaborative tasks*, and thereby would wholisitcally elicit and assess team schema-tasks-constraints via the mutual learning of team members, system designers, and other associated users that influence teamwork.

REFERENCES

1. U.S. Air Force Scientific Advisory Board (December, 1995). *New World Vistas: Air and Space Power for the 21st Century.* Information Technology Volume. Department of the Air Force.

2 Nosek, J., & McNeese, M. D. (1997). Augmenting group sensemaking in ill-defined, emerging situations. *Information Technology and People*, 10(3), 241-252.

3. Suchman, L. (1987). *Plans and situated actions: The problem of human-machine communication.* New York: Cambridge University Press.

4. Young, M. F., & McNeese, M. D. (1995). A situated cognition approach to problem solving. In P. Hancock, J. Flach, J. Caird, & K. Vincente (Eds.), *Local applications of the ecological approach to human-machine systems.* Hillsdale, NJ: Erlbaum.

5. Zsambok, C., & Klein, G. (1997). *Naturalistic decision making.* Hillsdale, NJ: Lawrence Erlbaum Associates.

6. Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, signs, and symbols, and other distinctions in human performance models. *IEEE Transaction on Systems, Man, and Cybernetics*, *13*, 257-266.

7. Gardner, H. (1985). *The mind's new science.* New York: Basic Books.

8. Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press.

9. Cannon-Bowers, J. & Salas, E. (1990, April). Cognitive psychology and team training: Shared mental models in complex systems. Paper presented to the Meeting of the

Society for Industrial/Organizational Psychology, Miami Beach, Florida.

10. Rentsch, J. R., & Hall, R. J. (1994a). Members of great teams think alike: A model of team effectiveness and schema similarity among team members. In M. M. Beyerlein & D. A. Johnson (Eds.), *Advances in interdisciplinary studies of work teams. Vol. 1. Series on self-managed work teams* (pp. 223-262). Greenwich, CT: JAI Press.

11. Rentsch, J. R., McNeese, M. D., Burnett, D. D., Bonnema, K. R. (1997). *The effects of individual differences and team processes on team member schema similarity and task performance.* (Final report prepared for Research and Development Laboratories and the Air Force Office of Scientific Research Summer Faculty Research Program). Wright-Patterson Air Force Base, Ohio: Armstrong Laboratory, Collaborative Systems Technology Branch.

12. Rentsch, J. R., & Hall, R. J. (April 9, 1994b). Person-environment fit mechanisms as regulators of optimal team member schema similarity. In B. Schneider (Chair), *Current thinking on person-environment fit.* Symposium presented to the Ninth Annual Conference of the Society for Industrial and Organizational Psychology, Nashville, Tennessee.

13. Steiner, I. D. (1972). *Group process and productivity.* New York: Academic Press.

14. Vicente, K. J. (1995). Task Analysis, cognitive task analysis, cognitive work analysis: What's the difference? *Proceedings of the Human Factors and Ergonomics Society*, 39, 534-537.

15. Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering.* New York: North Holland.

16. Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive systems engineering.* New York: Wiley.

17. Resnick, L. B. (1992). Shared cognition: Thinking as social practice. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition.* Washington DC: APA Press

18. McNeese, M. D. (1996). A comprehensive cognitive engineering process for human-information systems. *Proceedings of the 3rd Annual Symposium on Human interaction with Complex Systems, 1,* (pp. 260-267). IEEE Computer Press.

19. Sanderson, P. M., McNeese, M. D., & Zaff, B. S. (1994). Handling complex real world problems with two cognitive engineering tools: COGENT and MacSHAPA. *Behavior Research Methods, Instruments, & Computers, 17* (2), 117-124.

20. McNeese, M. D., Zaff, B. S., Brown, C. E, Citera, M., & Wellens, A. R. (1992). The role of a group-centered approach in the development of computer-supported collaborative design technologies. *Proceedings of the 36th Annual Meeting of the Human Factors Society, 2,* (pp. 867-871). Santa Monica, CA: Human Factors Society.

21. McNeese, M. D., Zaff, B. S., Citera, M., Brown, C. E., & Whitaker, R. (1995). AKADAM: Eliciting user knowledge to support participatory ergonomics. *International Journal of Industrial Ergonomics, 15,* 345-363.

22. Perusich, K., & McNeese, M. D. (1997). Using fuzzy cognitive maps to define the search space in problem solving. In G. Salvendy, M. Smith, & R. Koubek (Eds.), *Design of Computing Systems: Cognitive Considerations.* (pps. 805-809). Amsterdam, The Netherlands: Elsevier Science Publ. B.V.

23. Rentsch, J. R., Pape, L. P., & Brickman, N. M. (1997). *Team member schema similarity: Exploring individual differences predictors.* Unpublished manuscript. Wright State University.

24. Rentsch, J. R., McNeese, M. D., Pape, L. J., Burnett, D. D., Menard, D. M., Anesgart, M. (Submitted December, 1997). *Testing the effects of team processes on team member schema similarity and task performance: Examination of the team member schema similarity model.* (Technical report.) Prepared at Wright-Patterson Air Force Base, Ohio: Armstrong Laboratory, Collaborative Systems Technology Branch and Wright State University.

25. Boy, G. (1997). The group elicitation method for participatory design and usability testing. *Interactions* (March-April), 27-33.

26. Fisher, D. H. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, *2*, 139-172.

27. Baker, D. P., Salas, E., & Cannon-Bowers, J. (1998). Team task analysis: Lost but hopefully not forgotten. *The Industrial Psychologist*, 79-83

28. Bowers, C. A., Baker, D. P., Salas, E. (1994). Measuring the importance of teamwork: The reliability and validity of job/task analysis indices for team-training design. *Military Psychology*, 6, 205-214.

29. Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology*, *46*, 823-850.

30. Campion, M. A., Papper. E. M., & Medsker, G. J. (1996). Relations between work team characteristics and effectiveness: Replication and extension. *Personnel Psychology*, *49*, 429-452.

31. Schmidt, K. (1997). Cooperative work: Towards a conceptual foundation for CSCW Systems Design. *Human Computer Interaction* (pp.) In G. Salvendy, M. Smith, & R. Koubek (Eds.), *Design of computing systems: cognitive considerations.* (pps. 57-60). Amsterdam, The Netherlands: Elsevier Science Publ. B.V.

32. Wells, M. J., & Hoffman, H. G. (1997). *Proof of concept of the virtual pilot (VIP).* U. of Washington HITL Interim Report, 19 Dec 97.

33. Whitaker, R. D., Longinow, N. E., & McNeese, M. D. (1995). *Group interfaces: A profile and a prototype.* AL/CF-TR-1995-0069. Armstrong Laboratory, Wright-Patterson Air Force Base, OH.

34. Bannon, L. (1997). The role of common information spaces in CSCW. In G. Salvendy, M. Smith, & R. Koubek (Eds.), *Design of computing systems: cognitive considerations.* (pps. 317-320). Amsterdam, The Netherlands: Elsevier Science Publ. B.V.

Using Fuzzy Cognitive Maps to Assess Multi-Operator Situation Awareness

Karl Perusich, Ph.D.
Purdue University
1700 Mishawaka Ave.
South Bend, IN 46634, USA
kperusich@michianatoday.com

A key aspect of modern battlefield environments is the rapid generation and enormous volume of data available to decision makers. Although this data can potentially afford a decision maker an accurate "snap shot" of the emerging situation, it can also produce noise and data clutter that confuses or distracts. When the resulting operation requires individual members of a distributed team to independently make the correct, or, at a minimum, the same interpretation of the available data, volume can become a key impediment to achieving team situation awareness. The problems for decision making become even worse when the team members are not co-located and the responsibilities are distributed. To mitigate adverse effects and to coordinate decision making activities requires tools for developing shared situation awareness within the team. In this paper, the use of fuzzy cognitive maps to model the battlespace for developing shared situation awareness within a team will be discussed.

Team Situation Awareness

Situation awareness can be defined as the "...degree to which every team member possess the situation awareness required for his or her responsibilities." [1, p. 51] Situation awareness can be viewed at three different levels. At its most basic level (Level 1), a decision maker must perceive elements in the environment necessary for understanding the situation. This means that the technology, in a broad sense, is available to provide the "necessary" data and the decision maker has access to it. At a second level (Level 2), the decision maker must comprehend the current situation. He or she must be cognizant that a decision must be made, and how to implement that decision. At the highest cognitive level (Level 3), the decision maker must be able to project future states of different decisions to evaluate or identify one that best achieves the goals for the individual or team.

A team can be defined as having three characteristics: 1) a common goal, 2) interdependence, and 3) specific roles. It can be characterized as a group of decision makers with different responsibilities through whose coordinated action a common goal is achieved. Although the team members may share a common data base, through a common display or by being co-located in the environment, they typically use some subset of the available information for their decision making. If the decisions were completely independent relying on independent subsets of data there would be no problem. But in real team settings, decisions are characterized by overlapping data sets and feedback. Two or more decision makers may both require and use certain data. More importantly one person's decisions may be predicated upon what another person decides to do and vice versa, creating a feedback loop within the decision making structure. Achieving shared situation awareness with this team means that the members have access to the necessary information, including the decisions of the other team members, and that they are assessing it in a coordinated fashion.

In co-located teams, devices for achieving shared situation awareness are developed through three primary mechanisms.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

First, through direct communications team members indicate which data they are currently focusing on and what decisions they are undertaking. Direct communication can be achieved verbally, where literally one team members yells his or intentions at a particular moment, or it can be developed through non-verbal means such as watching what data another team member is focused on. A second device for achieving team situation awareness is shared displays. With a shared display the team members all have access to the same data and, hopefully, interpret it the same. In this way, individual actions undertaken work towards achieving the common team goal and are not incongruent. In a final means, team situation awareness is developed through embedding the co-located team in a common environment, for example, in the same airplane, so the team members each have access to the same environmental information.

A general trend in technology has been an increase in both the capabilities of data acquisition systems and of communications technology. This has lead to an increase in the use of distributed teams with distributed responsibilities that are not co-located. Achieving situation awareness in these settings becomes problematic because 2 of the 3 mechanisms just described are not possible. First, although team members can communicate verbally through a variety of technical means, direct communication as a shared situation awareness device is significantly diminished through the loss of non-verbal communication not easily achieved unless team members are in proximity to each other. Second, the distributed team members are not even necessarily on the same battlefield so a common environment is no longer a means for achieving shared situation awareness. It is now not uncommon for team members to be located in different states or on different continents.

The only remaining device for enabling situation awareness is through the shared displays and the information that is presented. The challenge for achieving shared situation awareness in distributed teams is to identify ways in which data and decisions can be presented to team members that insure they all interpret the emerging, dynamic battlespace environment the same and work towards achieving the same goal. And critical to this is the recognition of the feedback nature of the decision making process in the distributed teams. In the past, there has typically been a commander within the decision making structure in a military environment that has in some sense "imposed" situation awareness on a distributed team by interpreting the situation and providing the team members with the goals and tasks that they must accomplish. With the pace of a modern battlefield and the speed at which it evolves, having a central command act as a "clearinghouse" for developing situation awareness has become a bottleneck to rather than an enabler of team situation awareness. The real challenge to achieving situation awareness in distributed teams is to develop methods and mechanisms for sharing information, including information about decisions being made, so that all team members interpret the emerging battlespace the same and understand their role in achieving the common goal.

Fuzzy Cognitive Maps

To affect all three levels of situation awareness in a distributed team, a model is needed of the decision making process of all the team members that can incorporate a variety of data, information, decisions in the team setting and accurately reflect the tradeoffs necessary for identifying the optimal or best courses of action to achieve the team goal. Especially important, it must be able to model or effectively incorporate feedback within the decision making process. Such a model, to be truly useful, must be flexible, must be updatable in real time, and must provide the decision makers with a way to project changes in states for various choices or decisions they can make. This gives the team the ability to identify an optimal or best course of action to achieve the common goal.

In this paper, fuzzy cognitive maps will be examined as a means for modeling the decision making process in distributed teams with distributed tasks and distributed knowledge. A fuzzy cognitive map is a di-graph with fractional edge strengths that models the cause and effect relationships that a decision maker postulates about the solution space for a problem. [2,3,4] Each node in the map represents a cause or effect, with a directed edge between two nodes indicating a causal relationship. Edge strengths in the map are values on the interval [-1,1] with 1 indicating full causality (an increase in A causes an increase in B), -1 indicating inverse causality (an increase in A causes a decrease in B), and fractional values (negative or positive) indicating partial causality. This last case models situations such as "an increase in A *somewhat* causes an increase in B".

Nodes in a fuzzy cognitive map must represent variable quantities rather than instantiations of a particular concept. As such nodal values are restricted to the single values of -1 (decrease in value for concept), +1 (increase in value for concept), and 0, (no change in value for concept). A positive value represents an increase rather than a presence of the concept modeled by the node. Decisions (choices) are represented in the map by binary nodes, where 0 represents the absence of the decision and +1 represents it presence.

The fuzzy cognitive map can be used to infer the state of the system being modeled by the nodal values that occur after a set of "inputs" is applied and the map equilibrates. [8] A tradeoff is involved in designing and using a fuzzy cognitive map. Because concepts are represented in the map by nodal values rather than numerical quantities, a common numerical scale does not have to be developed for comparing conceptually different nodes. The map compares qualitative states of quantities to qualitative states of quantities, rather than numerical measures of concepts to numerical measures of concepts. This provides fuzzy cognitive maps with the distinct advantage of being easilty able to compare "apples to oranges". Very different types of concepts, such as environmental data, decisions being made, enemy intentions, and the status of weapons, can each be incorporated into the map and used in the inference process. The tradeoff, though, is that the fuzzy cognitive map can not determine numerical values for an underlying concept represented by a node. A map might infer that a *large* number of F-15's should be tasked, but it can not directly infer that 12 F-15's should be tasked. [7]

An important attribute of the fuzzy cognitive map is the ease with which several maps, independently constructed can be combined to produce a composite map that better reflects the overall situation than any map individually. Each individual map may only model a subset of the overall problem or reflect a portion of the available information. Since a fuzzy cognitive map compares states of attributes to states of attributes, individual maps can be combined by overlaying them. Common nodes are identified and used as anchors for combining the maps. Other nodes are simply added from these common nodes in such a way that the result reflects the connections identified in the individual maps. This ease of combining separate maps has great potential utility in using FCM techniques to construct an overall model of the situation that reflects the vantage points of the individual team members. The combined map can be used as a global or normative model to project future states, evaluate alternate courses of action and to identify potential conflicts in the decision processes of the individual team members.

Evaluating Fuzzy Cognitive Maps

To evaluate the efficacy of using fuzzy cognitive maps as a model for understanding the situation awareness of distributed teams with distributed information, a group of students was

divided into teams and assigned specific missions in a synthetic task. The task used was one of a hypothetical North Korean attack on South Korea in which SCUD missiles were used. Each student team was given responsibility for constructing rules of engagement for a particular asset that would likely be involved in any SCUD-hunting missions. For the scenario used, three teams, two of two members each and one of one member, participated. Each controlled (or tasked) one of the following air force assets: U-2 surveillance aircraft, F-15 Hunter/Killer groups or F-15's diverted from combat air patrol to attack a TEL.

The task was defined using the "Day After..." methodology developed at the RAND Corp. [9] This methodology provides the participants with a variety of background information pertinent to the problem, and has them walk though an evolving crisis. As they worked through the crisis they constructed a fuzzy cognitive map of the way that they would task the air force asset under their control. Although all three teams were provided with the same background information, they were not provided with the rules and models the other teams developed. The goal was to evaluate how well the individual models (as defined by a fuzzy cognitive map) would, when combined, produce an overall "team" model. The three resulting maps are presented in the figures at the end. A sequence of nodes in one of the map represents a chain of causal reasoning about how an environmental condition or a decision will affect other attributes. For example (from the map for Diverting F-15's from CAP), night would decrease the chances of finding the TEL which would decrease the ability to attack it. A decrease in the ability to attack the TEL would decrease the chances of destroying it which would decrease the desirability of diverting an F-15 from CAP to attack it. [5]

Results

Each of the three maps involved overlap. There were nodes common to two or all three maps, given in table 4. For example, the TEL fleeing was common to all three maps. Good and Bad weather was common to two. Not only were some nodes common but several of the models required input or knowledge of the decisions by another team. For example, tasking a Hunter/Killer group required knowledge of whether a U-2 had been tasked to locate and monitor the movements of the TEL. Likewise, the U-2 team required knowledge of whether an asset had been tasked that could attack the TEL should the U-2 be sent to monitor it. These nodal paths provided multiple feedback between the two maps that could affect the decisions made.

Combining the fuzzy cognitive maps produces a synthesized model of the decision making space that incorporates the attributes each of the team members identifies as important. This combined map can be used as a normative model or representation of the situation awareness of the team, i.e. it is a global snap shot of the battlespace from several different perspectives. It includes feedback and overlap. Identifying nodes and paths in the map that represent these two characteristics identifies both where the team *cooperates* to achieve goals, and where potential problems may exist. For example, there is no direct link between *Having a Sufficient Response Time* from the Tasking a Hunter/Killer Map and *Tasking a U-2* from the Tasking a U-2 map. But when the maps are combined an indirect path exists: *Sufficient Response Time → Sneak Attack → Multi-Front Attack → Tasking Multiple Hunter/Killer Groups → Attack TEL →Task U-2*. These hidden or indirect relationships that emerge only when the composite map is constructed represent the value added of a team effort over an individual effort. Additionally, they can represent significant constraints on the potential actions of one decision maker. Decisions by one team member can hamper or limit the available actions of another team, and not always in an obvious

way. Constructing a composite fuzzy cognitive map affords a vehicle for identification of these global relationships.

A composite fuzzy cognitive map can be used in several ways in the context of the situation awareness of a distributed team. Specifically:

- The map represents a global model or representation of the battlespace that incorporates decisions, environmental attributes, weapons capabilities, etc. This model can be used to understand the effect and the effectiveness of decisions.
- The model can be used to help different team members understand how their actions affect another's actions. A key feature of a composite map is that specialized knowledge is needed to construct it, but not to understand it. The map can be used as a communications bridge between team members with different knowledge.
- The model can be used as a software mediator for managing data and decisions. In past military epochs situation awareness was typically imposed by a centralized command and control structure that had access to all relevant data. This central command structure has now become a key bottleneck in the decision making process because of the increased pace of operations in a modern battlespace. A composite fuzzy cognitive map can be used as a software mediator to distribute data and determine optimal (or satisfactory) courses of action, and, consequently, foster a common understanding of the emerging situation among team members.
- A fuzzy cognitive map can be used as a diagnostic tool to identify potential incongrunencies in the decision making process of distributed teams with distributed knowledge. It is possible that if, left in isolation, a team member could evaluate the emerging situation from their vantage point, and make a decision orthogonal to the situation or the decisions of others. A composite map, because of feedback in the causal loops, may identify an alternate, more productive course of action for the team.

Using a composite fuzzy cognitive map as a diagnostic tool for identifying potential friction in team decision making will be evaluated in the context of the SCUD-hunting task used in this paper. The initial values for the scenario used are given in table 8. The values chosen were selected to try to give a situation where the U-2 surveillance aircraft would not be tasked, primarily because of the threat from enemy defenses. Applying these values to the *composite map* and inferring output values yielded the following result. A U-2 should not be tasked, nor should a Hunter/Killer group. The TEL flees and the chances of destroying it are reduced.

If the same initial conditions are applied to the individual map for tasking a U-2, the same results for the composite map are inferred. The U-2 should not be tasked. But if these initial conditions are applied to the individual map for tasking a Hunter/Killer group, a different result is inferred. Regardless of whether or not the U-2 is tasked, the inferred result from this map is that a Hunter/Killer group *should be* tasked, a result different from the projected decision from the composite map.

The composite map involves feedback, so certain effects are reinforced while others are dampened. This has the net result that the global conditions are such that a Hunter/Killer group should not be tasked. From the isolated view point of the Hunter/Killer group, as determined by its fuzzy cognitive map, a group should be tasked, but from a the more global viewpoint of the composite map, which includes two additional vantage points, it should not be tasked. Using a composite fuzzy cognitive map to identify and understand these discrepancies in decisions can be an important tool in improving the situation awareness of distributed teams and, consequently, team decision making.

Figure 1. Fuzzy Cognitive Map for Tasking F-15 Hunter/Killer Groups

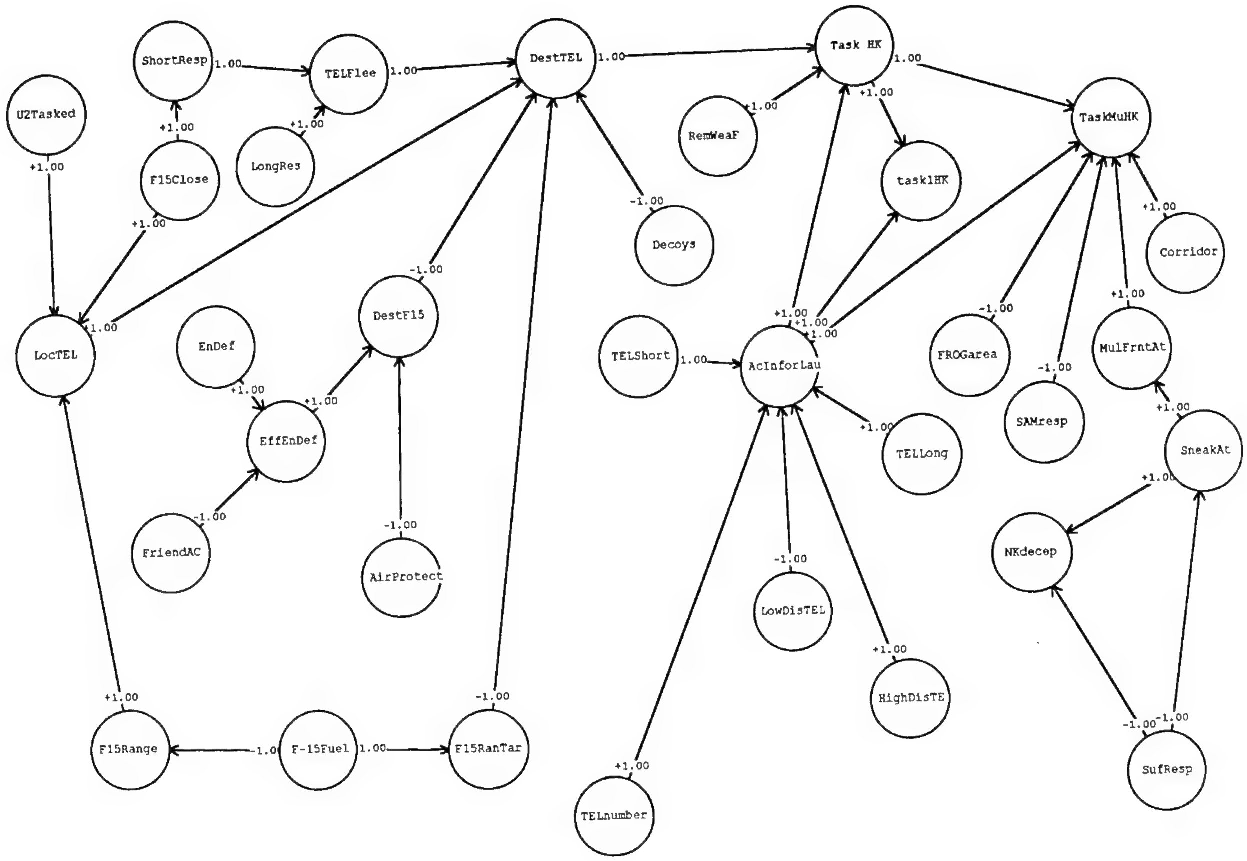


Table 1. Nodal Descriptions for Tasking Hunter/Killer Group Map

Node Mnemonic	*Description*
U2Tasked	U2 tasked
LocTEL	Ability to locate and monitor TEL
F15Range	Patrol range of F15
F-15Fuel	Fuel remaining for F15
F15RanTar	F15 range to target TEL
AirProtect	Availability of air protection for F-15
FriendAC	Other friendly aircraft in area
EffEnDef	Effectiveness of enemy defenses
EnDef	Enemy defenses in area
DestF15	Chances of destroying F15
F15Close	F15 close to TEL
LongRes	Long response time
ShortResp	Short response time
TELFlee	Chances of TEL fleeing
DestTEL	Chances of success of destroying TEL
Decoys	Presence of TEL decoys
RemWeaF	Remaining weapons load on F15
Task HK	Choose to task Hunter/Killer group
task1HK	Task one Hunter/Killer group
TELShort	Short coverage range of TEL's missiles
AcInforLau	Accuracy of information about launch area
TELLong	Long coverage range of TEL's missiles
LowDisTEL	Low accuracy in distinguishing decoys
HighDisTE	High accuracy in distinguishing decoys
TELnumber	Number of TEL's in area
SufResp	Sufficient response time for Patriot batteries
Nkdecep	Possibility of North Korean deception
SneakAt	Vulnerability to sneak attack
MulFrntAt	Possibility of multi-front attack
SAMresp	Response time for Patriot SAM battery
FROGarea	Coverage of FROG attack area
Corridor	Desire to create TEL-free corridor
TaskMuHK	Task Multiple Hunter/Killer groups

Table 2. Nodal Descriptions for Tasking U2 Map

Node Mnemonic	*Description*
TaskU2	Task U2
ImpCurMiss	Importance of mission U2 is currently on
AttackTEL	Availability of assets to attack TEL
Other U2	Availability of other U2's in theater
U2DesTEL	Chances of U2 being destroyed at TEL
U2Det	Detectability of U2 by enemy
HvEnDef	Heavy enemy defenses in route
LghtEnDef	Light enemy defenses in route
DesU2RT	Chances of U2 being destroyed before it reaches TEL
ShPatrol	Short patrol time for U2
LongPatrol	Long patrol time for U2
OverlapPA	Overlap of patrol areas
ShTransTm	Short time to reach TEL area
LnTransTm	Long time to reach TEL area
LtDefTEL	Light enemy defenses at TEL
HvyDefTE	Heavy enemy defenses at TEL
TELFlees	Chances that TEL can flee or evade
NumTEL	Number of TEL launchers in launch area
ImposTer	Impossible terrain to locate TEL in
FlatTerrai	flat terrain to locate TEL
DetectTEL	Detectability of TEL
BadWeath	Bad weather
GoodWeat	good weather
SensorUsa	U2 sensor usability

Figure 2. Fuzzy Cognitive Map for Tasking U2 Team

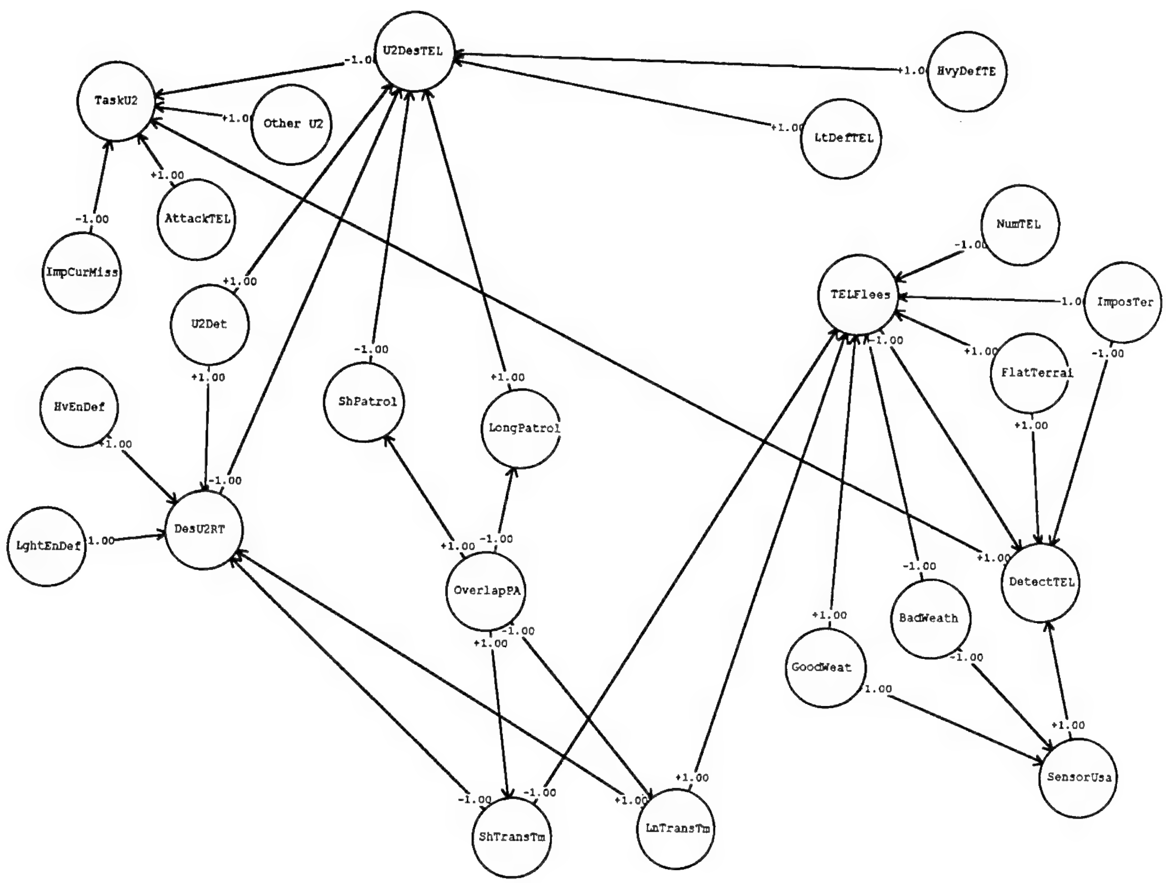


Figure 3. Fuzzy Cognitive Map for Tasking Other F-15 Assets

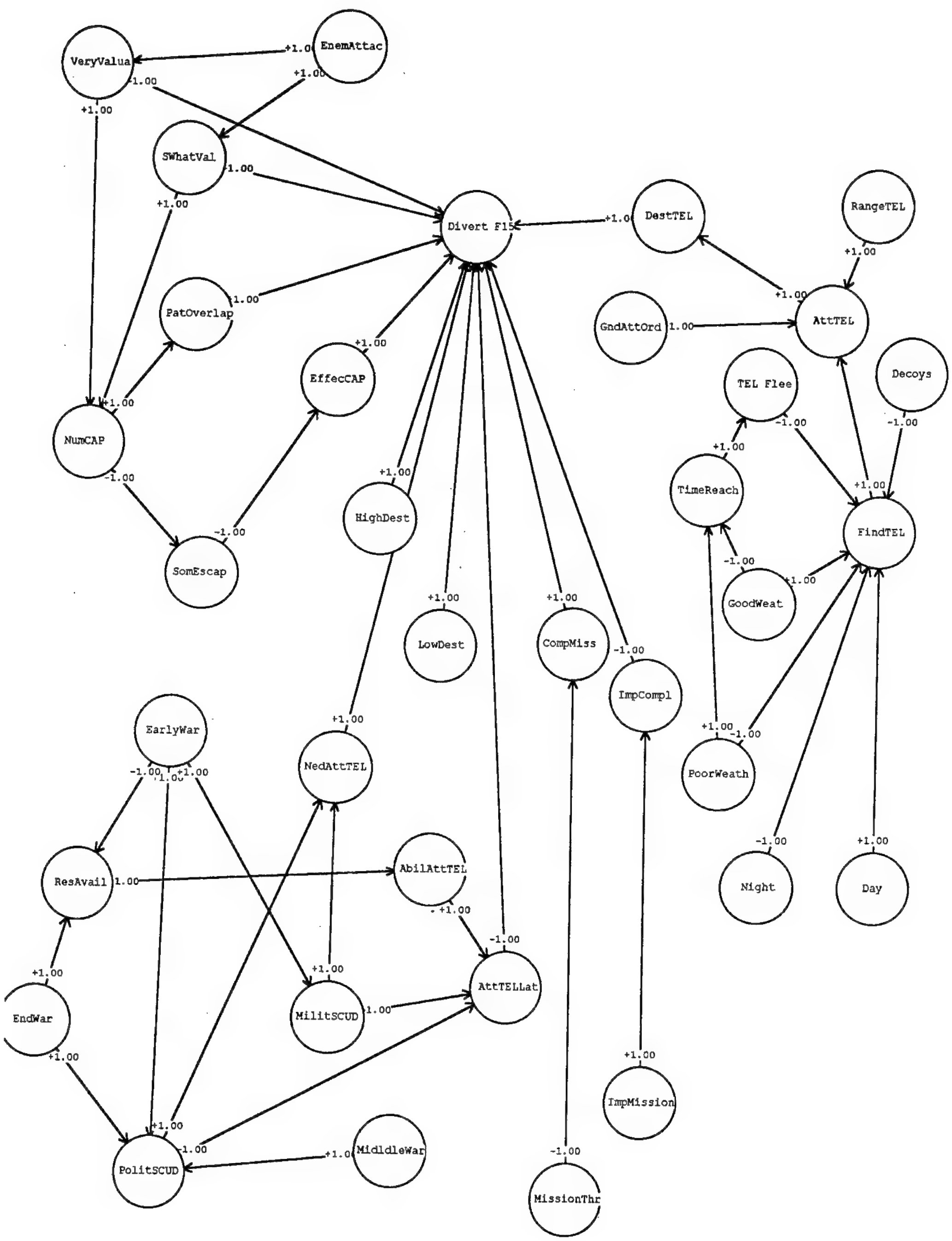


Table 3. Nodal Descriptions for Tasking Other F-15 Assets Map

Node Mnemonic	*Description*
VeryValua	Very valuable asset being protected
EnemAttac	Chances of enemy attack
SWhatVal	Somewhat valuable asset being protected
NumCAP	Number of aircraft on CAP
SomEscap	Chances of some enemy aircraft getting through
PatOverlap	Overlap of CAP patrol area with TEL launch area
EffecCAP	Effectiveness of CAP
Divert F15	Divert F-15's from CAP to attack TEL
High Dest	High value of destroying TEL
Low Dest	Low value of destroying TEL
CompMiss	Chances of completing mission
ImpCompl	Importance of completing mission
ImpMission	Importance of mission F-15 is escorting
MissionThr	Threat against mission
EarlyWar	Early stage of war
NedAttTEL	Need to attack TEL
ResAvail	Amount of resources available
AbilAttTEL	Ability to attack TEL later
AttTELLat	Luxury of attacking TEL's later
MilitSCUD	Military value of SCUD attack
EndWar	Near end of war
PolitSCUD	Political value of SCUD attack
MiddleWar	Middle stage of war
DestTEL	Chances of destroying TEL
RangeTEL	Range to TEL for F-15's
AttTEL	Ability to attack TEL
GndAttOrd	Ground attack ordinance
Decoys	Presence of decoys
TEL Flee	Ability of TEL to flee
TimeReach	Time of F-15 to reach the target
FindTEL	Ability to find and locate TEL
GoodWeat	Good weather
PoorWeat	Bad weather
Night	Time of day is night
Day	Time of day is day

Table 4. Nodes Common to Maps

	Map	
Tasking U2 Hunter/Killer Group	*Diverting F-15*	*Tasking*
Task U2		Task U2
Good Weather	Good Weather	
Bad Weather	Bad Weather	
TEL Flees	TEL Flees	TEL Flees
Number of TEL's		Number of TEL's
Enemy Defenses		Enemy Defenses
Task F-15	Task F-15	Task F-15
	Destroy TEL	Destroy TEL
	Presence of Decoys	Presence of Decoys
	Locating TEL	Locating TEL

Table 5. Inputs for Tasking Hunter/Killer Groups Map

U2 tasked
Fuel remaining for F15
Availability of air protection for F-15
Other friendly aircraft in area
Enemy defenses in area
F15 close to TEL
Long response time
Short response time
Presence of TEL decoys
Short coverage range of TEL's missiles
Long coverage range of TEL's missiles
Low accuracy in distinguishing decoys
High accuracy in distinguishing decoys
Number of TEL's in area
Sufficient response time for Patriot batteries
Response time for Patriot SAM battery
Coverage of FROG attack area
Desire to create TEL-free corridor

Table 6. Inputs for Tasking U2 Map

Importance of mission U2 is currently on
Availability of assets to attack TEL
Availability of other U2's in theater
Detectability of U2 by enemy
Heavy enemy defenses in route
Light enemy defenses in route
Overlap of patrol areas
Light enemy defenses at TEL
Heavy enemy defenses at TEL
Number of TEL launchers in launch area
Impossible terrain to locate TEL in
Flat terrain to locate TEL
Bad weather
good weather

Table 7. Inputs for Tasking other F-15 Assets Map

Very valuable asset being protected
Chances of enemy attack
Somewhat valuable asset being protected
High value of destroying TEL
Low value of destroying TEL
Importance of mission F-15 is escorting
Threat against mission
Early stage of war
Near end of war
Middle stage of war
Range to TEL for F-15's
Ground attack ordinance
Presence of decoys
Good weather
Bad weather
Time of day is night
Time of day is day

Table 8 Initial State Values for a Test Scenario for the Task

All nodes not listed below had initial values of 0.

Node Values of 1	***Node Values of -1***
EnemAttac	GndAttOrd
HighDest	OtherU2
RangeTEL	
Night	
ImposTerra	
LongResp	
ImpMission	
PoorWeath	
EarlyWar	
MissionThr	
HvDefTEL	
HKF15Clos	
AirProtect	

References

[1] M. Endsley and W. Jones. *Situation Awareness, Information Warfare and Information Dominance.* USAF Armstrong Laboratory. Technical Report 97-01. 1997. (Draft)

[2] B. Juliano and W. Bandler. *Tracing Chains of Thought.* Springer-Verlag. Heidelberg. Germany. 1996.

[3] B. Kosko. "Fuzzy Cognitive Maps." *International Journal of Man-Machine Studies.* 24. 1986. P.65-75.

[4] Kosko. "Adaptive Inference In Fuzzy Knowledge Networks". *IEEE International Conference on Neural Networks.* June. 1987. P.II-261-268.

[5] M. Mandeles, T. Hone and S. Terry. *Managing Command and Control in the Persian Gulf War. Praeger Press.* West Port. CT. 1996.

[6] K. Perusich. "Information Warfare: Radar in World War II as an Historical Example". *Proceedings of the International Symposium on Technology and Society 1997.* P. 92-99.

[7] K. Perusich and M. McNeese. "Using Fuzzy Cognitive Maps to Define the Search Space in Problem Solving". *Advances in Human Factors/Ergonomics. 21A Design of Computing Systems: Cognitive Considerations.* G. Salvendy, M. Smith and R. Koubek (Ed.). Elsevier. 1997. P.805-808.

[8] K. Perusich and M. McNeese. "Using Fuzzy Cognitive Maps for Data Abstraction and Synthesis in Decision Making". *Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society.* Syracuse. NY. 1997. P. 5-9.

[9] R. Molander, A. Riddle, and P. Wilson. *Strategic Information Warfare-A New Face of War.* Rand Corp. 1996.

Dysfunction Analysis
A Specification Method For Collective Work Situations ?

S. Hourlier
IMASSA-CERMA
Département de Sciences Cognitives
BP.73, 91223 Brétigny-sur-Orge Cedex
FRANCE
Sylvain@cerma.fr

SUMMARY

The growing part of work situations where crews have to deal with complex dynamic situations, justify the need for specifically developed assistance.The design of these systems relies on the supply of proper requirements to designers. An original method, relying on the analysis of an operator activity marker derived from breakdowns and referred to as "Dysfunctions", has been develloped. To assess its relevance in fulfilling the designer need, a study was conducted with data collected in the Operational Center of an Army Division. It was then presented to designers in charge of developing of a new Computerised Command System so as to evaluate its usability.
Results are presented with possible further developments

1. INTRODUCTION

This paper presents a complete methodology aimed at acquiring and processing human factors data in complex situations so as to give usable knowledge to designers confronted with designing assistance systems. Situations, be it collective, dynamic, involving risk management, with unavoidable deadlines... in a word "complex", are difficult and time consuming to analyse if any valuable data is to be collected. On the other hand, designers are in need of readily usable data on the human factors aspects of the work situations they are dealing with.

The developpement of assistance for operators working in complex situations has to rely on the usable knowledge of actual problems which these operators are facing, so as to address them in priority.
The situation used to validate this methodology is an Army Division Operation Center in combat exercise, for which a Communication and Information System is under development.

2. DYSFUNCTION ANALYSIS

2.1. JUSTIFICATION

The objective of this methodology is threefold. First acquiring data reflecting the operator needs and validating it with all those involved in the work situation (i.e. front line operators, hierarchy, domain specialists) to define priorities for design; second, being usable in any work situation, especially hard to access ones (military, operational,...) to get "in situ" data with no constraints on observees; last, realistically intergrating validated human factor data in a design process.

Design oriented human factors data can be obtained in various ways, ranging from single factor measurements to in depth observation with subsequent costly analysis. They can be collected on simulated environments, but nothing replaces the wholeness of real situations. Use of inservice experience is also a possibility, but as Branet & Trouilloud [1] put it, these reports carry inherent biases: they address mostly critical incidents linked to responsability aspects deterring their report, reports are *a posteriori*, thus highly subjective, they are felt like another constraint by operators, and last but not least, they rapidly become overwhelming for analysers due to their quantity. However, every method is limited by its ability to deliver usable data[1] when confronted with complex and hard to access work situations.

[1] for designers that is...

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

Dysfunction analysis draws it's originality from the analysis of specific activity[2] [2] markers, referred to as "Dysfunctions". There are two kinds of Dysfunctions: latent ones, and the ones emerging from the work situation and referred to as Stated-Dysfunctions (S-D). A S-D can be defined as:
«A spontaneous statement on the status of a Man-Machine System drifting from the functionning thought of as standard by the person(s) verbalising this statement».

S-D's are easily identifiable in real work situation, where they are the verbalised expression of the operator's inability to achieve expected objectives (either his or other's). As such, they reflect priority axis for development that designers should invest in, if they want to develop effective assistance for operators.

Dysfunctions are verbalised breakdowns. Breakdowns carry a dynamic aspect of changing state [3]. Identically, SDs are a consequences of a major alteration of an individual's planned schematas. Breakdowns are of high value as an evaluation tool, but verbalised breakdowns are of even higher value [4], [5]. Last S-Ds give access to L-Ds which open on a systemic approach to work situations.

For example, a typical S-D, emerged during assault (an extremelyactive stage in an Ops Center exercise). An operator temporarily alone in the intel division is trying to keep up with incoming radio reports from the regiments engaged in combat. Confronted to a generalised lack of accuracy in theses reports he finally says: "It'll never change.As soon as it gets hot, all reports are evasive and lack the proper coordinates to make them usable..."

These statements are situation specific and completely spontaneous, making them valuable and easy to collect. On the other hand, they are highly personal and subjective, because they refer to the individual's point of view. Such a bias has to be corrected by a cross validation with all other work situation participants.

This cross validation gives us the opportunity to unearth the Latent-Dysfunctions (LD) reponsible (in their conjunction) for the emergence of the Stated-Dysfunctions in the work situation. The front line operator is always the only one facing the consequences of the combination of various Latent Dysfunctions he is not responsible for most of the time. Their systemic character may take many aspects: faulty organisation, design error, non-synchronised decisions in higher spheres, obsolescent procedures in contradiction with environmental changes. For example, many SDs are linked to a LD that is the modification of Est-West relations and of the subsequent change in military deployement. Most of the procedures still based on predictible enemy doctrine proved to be somewhat ineffective during operations in the Gulf war and more recently in Bosnia. The LD here stems from the lack of reevaluation of the relevance of these procedures, rendered obsolescent by a changing world.

2.2. DESCRIPTION

The first step of the method is the collection of Stated-Dysfunctions during a work session. Though video is optimal if one has enough time to score it, a mere Pen & Paper method is sufficently accurate and pre-formats data, easing further processing. Anyway, S-Ds are quite easily recognisable in a collective work environment. The collection is twofold:
- a continuous detailed script of events, sufficiently rich to give operators the ability to grasp the situation in subsequent interviews,
- a concise, dated and indexed report of verbalisations corresponding to Stated-Dysfunctions.

The second step takes place during *a posteriori* interviews with the operators potentially involved in observed work situations:
- a cross validation is performed on each S–D,
- the S-Ds are placed in order in terms of importance,

[2] Activity is to be differentiated from the task. Task represent what the operator is prescribed to do and that he will have to interpret so as to execute it. Activity represents what is acually observed of what the operator really does in the work environment.

- L-Ds are attributed to every S-D,
- a cross validation of the L-Ds is also performed.

In the end each S-D is given a standard format containing: reference operator(s), referent situation, origin, date, comments made, rated importance and linked L-Ds.

3. EMPIRICAL STUDY

3.1. GOAL

An empirical study was set to field-test the method, and to verify the relevance of the data collected for designer use.
The work situation needed to be realistic, both complex and hard to access.
A new assistance system had also to be under design within the work situation being observed to assess the impact of the data on designers.
The questions addressed were data availability, quantity and representativity. Its usability was verified in an actual design process by analysing how designers perceived Latent-Dysfunctions behind each Stated-Dysfunction (as compared to operators) and what solutions they provided.

3.2. SITUATION

The situation was the yearly field exercise of an Army Division Ops Center in May 97. The observations were made over three days by a single human factor specialist in the Intelligence Section of the Ops Center. This section is maned by two shifts of 5 intel specialists, and runs 24 hours a day.

4. RESULTS

During three 8 hours observation sessions, 26 Stated-Dysfunctions in the Intelligence Section of the Ops Center were collected.The testing of the collection part of the method was a definite success considered the poor local support obtained at the time.They were definitely doing their jobs and were not there for any other reason.

During *aposteriori* interviews, 25 Stated-Dysfunctions were validated, and only one was rejected, because it disagreed greatly with the intel experts present at the time. The operators potentially involved in the observed work situation attributed these Stated-Dysfunctions to a total of 25 Latent-Dysfunctions. The Stated-Dysfunctions being related to an average of 2 or 3 Latent-Dysfunctions.
Here is a list of the Latent-Dysfunctions revealed during the interviews:

General organisational problems (27):
- lack of personnel (12),
- lack of room (6),
- unforseen variation of task demands (9)

Communication problem (sharing of SA) (1)

Training problem (9):
- procedure disrespect (2),
- means and time to train (1),
- lack of common language (1),
- reading and writing of intel (3),
- planification of intel prospection (2),
- individual capacities (1),

Tool/interface adaptation problems (9):
- unadapted to collective work aspects (1),
- unadapted to use evolution, no replay possible (1),
- difficulty to recover when interupted (1),
- unadapted to type of data (1),
- goal conflicts (1),
- presentation of data lacking visibility (3),
- network insufficency (1),

Local organisationnel problems (8):
- allocation of ressources (5),
- priority conflicts (1),
- lack of intel acquisition system (1),
- allocation of room / need (1),

Procedures adaptation problems (4):
- lack of procedures / evolution of task demand (1),
- lack of procedures / hierarchical contersigning (1),
- sterile mesages steam due to lack of delegation (1),
- inadaptation to new warfare context (1).

The last part of the study was devoted to analysing the designer's attitude towards S-D. Two designers involved in the development of an assistance for the Ops Center operators were presented the 25 S-Ds. They were asked to mention the possible Latent-Dysfunctions they were connected with, an to present how they would take them into account in design.

The connections made by designers (Figure 1) were the same as those of the operators in more than half the S-Ds, though no connections were made in one fifth, the rest being partial connections (involving only the technological aspects of the S-Ds).

Three designer attitudes could be identified:
- not concerned with the S-D (7/25). For most of these S-Ds they identified non-technological L-Ds,
- taking into account only the technological aspects of the S-D (11/25), whatever L-Ds they found out, and suggest an "off the

shelf" solution, incompletely dealing with the need,
- taking into account the S-D with all their connected L-Ds (7/25) and, not finding any readily usable ("off the shelf") solution, proposing a total reformulation of the problem in specification terms usable by designers.

LD Connected to SD				Attitude of designers	
Same	Partial	None			
3		4	7	Not a design problem	A
5	6		11	Only technical Pb addressed	B
7			7	Technical an HF problems addressed	C
15	6	4	25		

Figure 1: Connections made by designers and solutions proposed; A=None, B=Partial, C=Reformulation

A typical example of this last attitude is the suggestion made to S-D N°6 (where operators became crazy during a particularly tense moment with outside operators interrupting them constantly for situation updates). For this S-D, the designers we interviewed, having no "off the shelf" solution, creatively suggested to build a device that would "give the interrupting people what they may need before they interrupt". Of course this is not a solution yet, but this kind of formulation denotes a good grasp of the wholeness/complexity of the factors involved and as such accounts for all the aspects of the problem, and not only the technological ones.

5. DISCUSSION

As presented on Figure 1, it is easy to perceive the advantages of presenting S-Ds with their connected L-Ds and to have them explained by a human factor specialist

- the horizontal arrow would be the effect of of data enrichment (i.e. presenting also the connected L-Ds)
- the vertical arrow would be the effect of a human factor specialist support to enhance the HF aspects needing to be addressed in a possible solution.

This model of designer attitudes outlined throughout their use of Stated-Dysfunctions needs to be comforted by further studies.

6. CONCLUSION

The field study gave the opportunity to assess the value of the dysfunction analysis method :
- as a tool for complex and hard to access work situations,
- for collecting HF data relevant to operator needs in terms of assistance,
- giving access to a large panel of events and constraints ranging from the individual's work situation to the whole system.
- and finally producing usable data for designers, as long as they are supported by HF specialists.

BIBLIOGRAPHY

1. Branet, C., and Trouilloud, J., (1996). "Intégration du retour d'expérience non-statistique dans le processus de conception". *Actes du 10ème colloque national de fiabilité et maintenabilité,* Saint-Malo, (pp. 397-411).

2. Leplat, J., and Hoc, J. M., (1983). "Tâche et activité dans l'analyse psychologique des situations". Cahiers de psychologie, 3(1), 49-63.

3. Bødker, S., (1989). "A Human Activity Approach to User Interfaces". Human Computer Interaction, 4, 171-195.

4. Urquijo, S. P., Scrivener, S. A. R., and Palmén, H. K., (1994). "The use of breakdown analysis in synchronous CSCW system design" (background paper).

5. Wright, P. C., and Monk, A. F., (1989). "Evaluation for design". In A. S. L. Macaulay (Ed.), People and Computers V., . Cambridge, UK: Cambridge University Press.

Communication Requirements in the Cockpit

Dr Malcolm J. Cook and Dr L. Elder
Division of Psychology
School of Social and Health Sciences
University of Abertay Dundee
158 Marketgait
DUNDEE, DD1 1NJ, UK
Tel. 01382 308749 - Fx: 01382 223121
Answerphone : 01382 308709

George Ward
ESE Associates
15, Jesse Close
Yately, Hampshire GU46 6AH, UK

e-mail : m.cook@river.tay.ac.uk
George.Ward@bae.co.uk

Summary

Communication is a core activity in support of aeronautical decision making in both civil and military aircraft. Decision making in groups occurs through interpersonal communication based on exchange of information (De Sanctis and Gallupe, 1987). It has been recognised that free exchange of information between man and machine is no less important than that between the human operators within cockpits, pilots and air traffic controllers, or between pilots in different aircraft. It is also unlikely that effective communication between human operators and machines can achieve as much in terms of improving performance as the communication between human operators for the simple reason that human operators are more context sensitive and integrate a wider body of information in their decision making (Mosier and Skitka, 1996). Thus, review of actions, intent, plans and goals by human operators represents a potentially richer source of critiscism than that afforded by computer-based critiquing systems. If one accepts this view then one accepts the need to protect communication between humans from interference, disruption or enforced reduction because this would jeopardize the quality of decision making. In turn reduced quality or frequency of communication could compromise the effective performance of the individual operators and the coordinated response of teams of operators.

This paper presents evidence in support of the view that communication requires careful assessment, and the introduction of new technologies must be carefully assessed to address the changes they may produce in communication patterns. One of the reasons why communication is more likely to be subject to changes in systems in multi-crew multi-platform or multi-crew single platform systems is the role of sensitivity of collaborative systems to disf-engagement. In simple terms it has been recognised that collaborative applications and systems require multiple users if they are going to be successful. If users feel that the communication tasks interfere with other functions or are difficult to use then multi-user systems will fail (Grudin, 1988). There is ample evidence from civil aircraft cockpits that unwanted and damaging effects can be produced by new automation which is not directly involved in communicative processes. It is, therefore, reasonable to assert that systems which directly interfere with communicative processes need careful assessment because of the unintended effects on free exchange of information on intent, goals, plans, actions and outcomes.

Introduction

This paper is intended as an integrative view of the possible effects of new technologies in civil and military cockpits and their likely effects on interaction between operators and concomitant effects on both team and individual performance. Much of the work relies upon the foresight of workers in the field of aeronautical human factors (Bowers, Oser, Salas and Cannon-Bowers, 1996) but some of the views have been extracted from the rapidly developing field of computer-supported cooperative working, in particular Mantovani (1996). This paper aims to relate the concerns expressed in those areas to the importance of communication and the likely effects of changes introduced by new technologies. It is suggested that designers recognize that technologies both directly and indirectly concerned with communication can have significant effects on interaction, which can in turn can lead to changes in performance. Analysis of research literature seems to indicate that designers of equipment do not consider the secondary effects of communication in the design process despite recognition of the importance of communication in distributed decision making (Brehmer, 1991).

The Role of Communication in Future Systems

It is possible that reduced emphasis will be put upon communication skills and communication systems in the near future because of a number of changes which are likely to take place in future systems. Communication is already recognised as a resource intensive activity for two-seat military operations and it is likely to be more demanding in future single-seat operations. It is a simple fact that managing communication in time within single or multi-platform multi-operator systems is problematic. Communication activities could easily detract from other safety and mission critical tasks. Tactically communication may not be an option for first look - first kill type engagements where the operator attempts to maintain a stealthy aircraft for as long as possible. The danger is that this might be translated into a policy for a reduced requirement in training communication and coordination skills in tandem. At present there are very limited facilities for inexpensive multi-operator training that do not involve ranges and which can not be used aggressively without risk. Recognising the increasing costs of maintaining operators at their peak level of skill it is clear that an advanced simulator system for training combat weapon skills and manouevring would be advantageous.

The next question to address is the level at which communication should take place. Many systems present raw information or offer limited abilities to manipulate the incoming streams of data. It is

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

almost universally agreed that pilots of military aircraft have an unenviable task of integrating a diverse set of data streams in order to take effective decisions about actions and their potential consequences. Offering manual filtering of incoming data streams reduces the flow but at the cost of introducing another task for the pilot to perform. Removing the pilot from the data processing activities creates disengagement from the tasks and increases the possibilities for faulty situational awareness. Similarly with communication there are options for passive and active communication strategies. Either the pilot assumes a stealthy option in which directions are given by air controllers or package managers on the ground or in the air, or, the pilot plays an active part negotiating their role in the mission scenario developed. There is clearly a third option for disengaged opportunistic activity and a military equivalent of free-flight. This third option is not likely to be operationally effective and may create significant risks.

Information Sharing and Shared Meaning

A fundamental premise of this work is the difficult problem of shared meaning in distributed systems because some authors have taken the extreme view that independent agents may not have identical representations, as a result of differences in prior experience and differences in the operators' local environment (Gasser, 1991). The question of meaning is not a moot point because it depends upon the level at which teams negotiate future action, plans or goals. It is quite possible that negotiations take a simple form which is required for the alignment of actions and the reduction of conflict. Alternatively, the interactions may implicitly assume a mental model and negotiation is built on presumed shared assumptions. Even where mental models exist it is generally agreed that they are incomplete, unstable, and parsimonious which means that people try to maintain them by communication or action (Boehm-Davis, 1990). Whether one accepts or rejects the view that shared mental models are possible the alignment of contributions through coordination is a necessary feature of distributed teams which in turn requires a careful analysis of the communication opportunities.

Simple forms of interaction may make fewer assumptions and result in more protracted communication styles because an agreement must be developed. Many operators do not immediately recognise that negotiation is often taking place, albeit, in a covert manner because it is only apparent when challenges are made to requests for faulty or inappropriate action. The protracted style of negotiation would be subject to influences of time pressure and workload with performance decrements more likely as operators sacrifice time to maintain communication at the expense of other activities. More sophisticated forms of interaction may be more succinct but they may be prone to error if individual operators activate an inappropriate model. Recent evidence suggest that RAF strategies are directed towards effective combat weapons system training where emphasis is on practical experience and the development of skill-based performance supported by knowledge gained in exercises (Barrie, 1998). This, in turn, would indicate a preference for the development of an effective shared mental model and less communication, which is largely focused on management of mission related tasks in time and space. It is debatable whether it is always possible to maintain the level of intelligence required or make preparations to support that level of exchange among the operators in a multi-ship engagement.

Indeed, recent evidence suggests that decision making in response to relatively well understood procedures can still fail to produce non-optimal results when the communication skills and acts of the participants are ineffectual. For example, Spiller (1997) reported that after a collision between two Tornados in transit across Canada the communication between the pilots and those in the other aircraft, a Lockeed Tristar and those in the other two Tornados present, was ineffective in developing an effective response to the unfortunate accident. Similar problems have been observed and reported in civil aircraft operations with failure to communicate the severity of fuel shortage and aircraft status contributing to an accident when an aircraft ran out of fuel (Billings, 1997). In another case the aircrew, Captain, First Officer and Flight Engineer failed to notice that they had disengaged the autopilot and as they descended into the ground the air-traffic controller asked " Is everything alright up there ?" and did not query the specific change in flight status of the aircraft.

Some of the ideas presented here were developed and presented in a paper at an IEE Colloquium in 1997 on Computer-Supported Decision Making (Cook, Elder and Ward, 1997a, 1997b) and by Cook (1997) in which it was suggested that information sharing is not equivalent to sharing knowledge because knowledge implies interpretation which may not be common to all operators working with the same information. It was suggested that effective communication was a vital part of the process whereby agreed solutions and action-plans might be generated. Beach (1990), has suggested that 'much of the social interaction that precedes group decision making is devoted to ironing out differences in the participants' frames through the sharing of information and through negotiation'. Thus, it seems obvious that disrupting the process of negotiation would in turn have potential effects on the outcomes of distributed operators performance. Disruption of communication has always been a prime directive in any conflict because of the potential effect on the opposition's coordinated response, there is no reason to expect that it operates only at the highest level.

Some have suggested that members acting cooperatively must use a shared system of meaning. For example, Trevino et al. (1990) has suggested that "People have developed many shared assumptions and understandings about the meanings of words, actions and events". It would be trite to recognize the importance of training in forging a common system of interpretation but this would fail to recognize the much more liberal approach to tactical decisions, and presumably decision making, encouraged in operational conversion units of some air forces.

Table 1 - Inserts here - See appendices.

It is important to recognize that a balance must be struck between order and flexibility in decision making during dynamic situations. Restricted information flow, limited consideration of options, very formal communication within the group and no open disagreement can be the conditions which are conducive to Groupthink (Janis, 1982; Eiser 1986; Hartley, 1997). Given that the decisions are made under high stress, decisions could be made with a lack of considered alternatives and a failure to develop contingency plans producing the conditions which are a perfect breeding ground for faulty decision making. Opening up the discussion to the consideration of alternatives in pre-briefing provides flexibility

but it must be managed dynamically by inter-operator communication during the missions. It is clear that decision making is prone to err but the social psychological studies make clear that the communication which supports effective decision making and action is in turn supported by experience which underlines the importance of training communication skills beyond the level of task-related vocabulary.

It has been suggested that units working with the same method of sharing information, JTIDS, have independently developed two action strategies for using the information afforded by the communication system used. This flexible implementation of technological usage is not the orderly management of information flow envisaged by the technologist and it reflects the view that -

"In reality, what often happens when new technologies are introduced into cooperative activity is not simplification but greater complications, because the new technologies are far from being crystal clear to their users." Mantovani (1996, p. 86-87).

Informal discussion suggests that the training regimes and views put forward by central training establishments may even differ from that encouraged by operational units. The RAF appears to be taking significant steps to rectify the differences in the approaches to training in recent years and synthesizing a new approach by encouraging communication between Operational Conversion Units (OCUs) and the Air Warfare Centre (AWC). Even so the application of technology and their integration into the imagined roles may be extremely fluid. Even if the strong form of the observations with regard to differences in training by OCU and AWC are no longer tenable it is clear that an intermediate form may be necessitated by exigencies of combat because that flexibility is required to maintain operational effectiveness in unpredictable circumstances via a range of possible actions. Whatever the conclusions, it is clear that pre-planned responses cannot be a total solution because -

" It is impossible both in practice and in theory to anticipate and foresee every contingency that could occur during the development of a set of tasks. Therefore, no formal description of a system can be complete" (Gerson and Star, 1986).

This quotation is supported by the observations from real-life noted by Spiller (1997) and indicates the need for distributed decision making in multi-platform multi-operator teams. Similar views have been espoused by Brehmer (1991) who has argued that the review of the literature underlines what practical experience suggests that communication is costly and seems to encourage inefficiency. However,

"If a system is complex, some form of distributed control is necessary, and if all problems that the system must face cannot be foreseen, communication is needed for the system to reconfigure itself. It is necessary to not introduce organisations that make such reconfiguration impossible." (Brehmer, 1991, p. 12).

Accepting the ambiguity in interpretation of unfolding events revealed by information, the unpredictable nature of environmental and contextual demands, and the variety of roles afforded the participants is central tenet of the thesis of this paper. If the interpretation of the world was unambiguous and the required actions obvious there would be no need for communication during cooperation. It is no more the case in air-defence than in other forms of mediated communication as Mantovani (1996) points out.

"We say *cooperation* for the sake of brevity, but it should not be forgotten that cooperation, conflict and negotiation are all inextricably intertwined (Mantovani,1996, p. 83).

Operators or in the case of aerospace systems, pilots, can be taught the principles of analysis but they must develop the interpretive skills for understanding the information through experience. Exercises like Red Flag must contribute to developing skills in analysis and interpretation of information. Even so it is not clear that operators will experience the full range of possible situations which might present themselves in the operational history of the pilot. Nor is it clear that operators will use the full range of knowledge in their action choices for knowledge based reasons or inherent biases in processing (Beach and Mitchell, 1990; Beach, 1991; Wickens and Huey, 1993). It has been suggested that within teams of distributed operators one can distinguish two levels of information processing : the primary mode in which the individual's cognition is involved; and a secondary mode in which intra-group communication takes place (Leplat, 1991). If these processes are tightly coupled through hierarchical organisation or limited communication opportunities it is clearly possible for faulty decision making to undermine total team performance. If intra-group communication is more open then critical reviews of proposed actions, coordination or demand evaluation can occur.

Trouble with Technology

Recognition of the potential differences in technology usage seems to run counter to the view that uniformity of outlook is developed in training and that situational ambiguity is reduced. Bowers et al. (1996) recognised that automation tends to increase the complexity of tasks and the level of demand and they were not alone in recognising the possibility of increased workload after the introduction of automation (Mosier and Skitka, 1996; Parasuraman, Mouloua, Molloy and Hillburn, 1996; Kantowitz and Campbell, 1996). Increased workload is likely to result in task-shedding at high levels of demand and inter-operator communication is likely to be one of the first areas to evidence change. Informal discussions with crash investigators indicates that linguistic silence frequently occurs in the final portion of cockpit voice recorders because pilots are devoting all of their resources to their attempted recovery of the aircraft. Communication technologies in office and work environments may seem to be very different in their requirements to the types of system operated in aircraft but the common ground is the unpredictable nature of the effects of the introduction of new technologies. Representative quotes from these areas can be strikingly similar :-

"The current disorientation regarding the social and cultural implications of technological innovation reveals not so much a presumed immaturity of new technologies and their experimental character, as is often said, but rather the absence of appropriate cultural responses to questions posed by the current forms of technology-organisation mix." Mantovani (1996, p. 79)

In discussing the role of behavioural scientists in the assessment of such team-oriented military systems it has been said that there is :-

"little opportunity to prevent the unforeseen, potentially catastrophic consequences of poor performance in these systems. Rather behavioural scientists will be remanded to the rather unsatisfying duty of attempting to understand why the performance deficit occurred". (Bowers et al., 1996, p. 258).

Both of these quotations indicate that predictions of technological impact on performance are problematic and prone to mis-representation of expected outcomes, after the introduction of new technologies. This is all the more alarming when one considers that new technologies are being introduced into military systems when there is increased pressure to reduce crew complements and change the roles of the participants. The broad similarities of this pattern of changes is all the more disturbing if one considers that this is essentially the scenario that occurred in civil cockpits when the flight engineers were taken out of the cockpit and significant increases in system automation occurred. It can be argued that many of the potential problems observed in those changes on civil flight decks could quite easily occur in military aircraft and give rise to similar outcomes. Indeed, some operational pilots have expressed the view that an increase in the attrition rate of pilots is to be expected when new aircraft are introduced because of need to accommodate changes in roles and equipment.

The selection of single seat solutions for potential strike and air defence aircraft seems all the more surprising when one considers the opinions afforded by some authors :-

" The effectiveness of two-man crew was graphically shown in the results achieved by *two holers* like the F-111, A-6 and the F14E. With defensive systems getting more and more effective, it will take an adept mission to prosecute the mission tasking and survive to fight another day." (Isby, 1997, p170).

It may be the case that combination of communication and sharing of workload afforded by two-man crews increases the total effectiveness of both the single aircraft and the coordination of all the aircraft involved in an engagement. Thus, it is very easy to dismiss the role of communication in supporting multi-operator activities because it is largely encapsulated within the cockpit in current two-seat aircraft.

The concern expressed with regard to future technologies and future changes in roles confirms the view proposed by Bowers et al. (1996) that technological issues are often considered first and the emphasis during technology development is largely on what can be achieved. In addition, the operational costs of two operators, pilot and navigator, are not far behind and it is hard not to believe that this encourages the shift towards unmanned air combat vehicles. Again the performance issues raised by the design decisions taken for new systems are often considered after construction of the new artefact or technology when changes are too expensive to implement. For example, a personal communication with a vendor supporting the Joint Tactical Information Distribution System at a trade conference revealed that the integration issues of supporting communication systems parallel to JTIDS had largely been ignored or it had been implicitly assumed that these would be dealt with by the system integrator or the operator.

Mosier and Skitka (1996) noted that automation had resulted in changes in the pattern of communication in respect of the decision making process in the civil flight deck because information flow had ceased in some cases or changed in a qualitative, quantitative or combined way. The Glass Cockpit as it is now known had subtly changed crew coordination and interaction, with crew members seeking information from the automated systems rather than each other. It is interesting to consider the parallels between the Glass Cocpkit and the provision of decision support and shared radar images in future military cockpits. In civil cockpits awareness of others knowledge, plans, intentions and goals declined in response to the new technologies. At the same time the workload and the roles of the operators changed with the dissappearance of the flight engineer. Thus, changes in technology and systems interfaces were overlaid on top of new functional requirements on the operators. It is clear that the changing pattern of information flow is critical in establishing the roles of participants in a dynamic situation because -

" if information flows are changed, the new technologies will alter the terms and modes of social interaction". Mantovani (1996, p. 87).

Changes in communication have been noted in automated cockpits, compared to standard cockpits, with changes more marked in high workload environments (Bowers et al., 1996). In addition it has been found that crews composed of individuals with heterogeneous levels of experience obtain greater benefit from automation than those in homogeneous crew complements (Wickens et al., 1989) and subsequent analysis of heterogeneous crews' communication indicated an improvement in communication (Straus and Cooper, 1989). It is not clear that the communication in teams with differing levels of experience operating in an highly-automated environment were not subjected to greater role and behavioural uncertainty which encouraged more effective communication. Thus, in highly experienced crews it might be expected that a certain degree of automation induced complacency could occur and reduced communication would result in negative consequences given the appropriate circumstances. Suggestions of negligent behaviour and complacency of crew in civil and military accidents have been more frequent in recent years but it would be difficult to establish changes in behaviour encouraged by automation as the root cause of that increase.

Another alternative conclusion that might be drawn from research on communication among crews with heterogeneous experience and automation is the need for communication to resolve ambiguities in roles, situations or actions. Thus, artificial suppression of communication through standard operating procedures or reduction in communication imposed or encouraged by automation might result in performance decrements due to a reduced quality of situational awareness. Military operational engagements may dictate that crews with different levels of experience are required to operate together and denying or reducing effective communication in those circumstances may precipitate a greater frequency of accidents, incidents or mission failures. If communication systems are not effective or they are disrupted by the demands of on-board equipment then individual users may fail to adopt the technology for use and a collaborative system fails.

The Functions of Communication

Malin et al. (1991) and Malin and Schreckenghost (1992) have identified communication as a key activity in coordination and collaboration activities between human-human and human-machine operators. In an analysis of organisation of collective tasks Leplat (1991) has suggested that communication plays a vital role in ensuring the successful completion of collaborative tasks. It has generally been accepted that management and scheduling of the dialogue between different operators or operators and machines is an additional task. As a consequence communication tends to be a very demanding task (Brehmer, 1991). Essentially there is a requirement to transmit information and manage the transmission which means that it may suffer as a result of task-shedding in high workload phases of a task. Where task shedding of communication does not occur there are suggestions that it may interfere with other activities that the operator is carrying out. For example, there is anecdotal evidence that mobile phones used in moving cars may divert sufficient attention away from the visual field to increase the likelihood of accidents and experimental evidence reported by Cook M.J., et al. (1997c) and Cranmer and Cook (1997) indicates that strong interference may occur between high demand visual vigilance tasks and direct voice input and output systems. Indeed, there is some evidence that it is the coordination of communication between participants and not the receipt of information per se that may be the most demanding aspect of intra-group interaction. However, the quality of the communication and its value may be reduced by simple broadcasting of information between participants because this has been an inffective strategy with intelligent agents.

Table 2 - Inserts here - See appendices.

Two roles for communication have been identified by Malin (1991) and Malin and Schreckenghost (1992). First, communication can be used to repair team processes because action schemas may need to be substituted or modified based on new information which is disseminated by communication. Second, communication is required to ensure dynamic coordination of action because optimal performance may only be achieved in multi-operator activities by temporal and spatial coordination of actions. Both military and civil cockpits can use communication to perform information dissemination and action coordination functions. It has been suggested that communication may not always inform people but it may assist in the pragmatic alignment of activities and action as if shared knowledge was available. In simple terms, communication may in some cases only help to reduce conflicting activities by ensuring actions are distinct but it may not create successful parallel actions which are coordinated.

Effective communication occurs when information is distributed when needed and in a form which can be assimilated quickly. It is not clear that this has been carefully considered in the current design process. Often the strategy has been to present a great deal of information, continuously and in a raw form which does not answer the kinds of questions pilots want to answer. Only latterly have filtering schemes been applied to information display in airborne early warning aircraft and in single-seat operation even these forms of enhanced display format may be offered in a much less sophisticated form. The issue of temporal relevance has largely been omitted because of the problems in providing robust images of the sensed environment and the limited intelligence on-board aircraft.

Woods (1994) has suggested that automated agents could assist in coordinating activities among human and automatic agents by assisting in the process of sharing information. The only danger in this is the appearance of presumptive decision making based on partial information and care would be needed to ensure that training addresses this possibility. Even after training it could be the case that presumptive decision making occurs under high stress and high workload conditions which would indicate that management of information sharing is as important as the act itself. To adequately address the functions of communication and the different types of communication it would be useful to address the goals which are to be met by communicative processes.

Goals of Communication

Communication can largely be said to influence decision making through the development of a mental model which in turn influences decision making. A similar point view has been put forward for automated systems and decision support systems some time ago (Silverman, 1992). Automated systems had failed to achieve the levels of communicative sophistication or context sensitivity to support these modes of decision support because they have little flexibility.

A model of decision making proposed for Civil Aircraft Flight Decks (Donnelly, 1996; 1997) can be adapted for use in analysing decision making in military aircraft and two points of intervention for improving decision making via communication are revealed. First, at the presentation of information and the effects are largely in terms of improved situational awareness. The second point of intervention is prior to emission of an action or before the action takes full effect and this communication would take the form of a challenge to the action. The proximal goal of communication in both challenges to proposed action and in furnishing general information is matching of available resources to situational demands. This assumes that interpretation of these basic facts needs no further explanation and this may be true in certain situations.

Table 3 - Inserts here - See appendices.

The seven categories of team function are orientation, resource distribution, timing, response coordination, systems monitoring, motivation and procedural maintenance. Many of these functions are clearly important in dynamic situations with timing, response coordination and system monitoring functions among the most obvious (Leplat, 1991). However, functions like orientation and motivation can be very important in situations where transitions in context or demand are very high and multiple operators need to coordinate an effective response to the changing demand.

The situational demands and the resources available to tackle them can be grouped under 4 headings (Nieva, Fleishman and Rieck,

1978) : resources, task characteristics, team characteristics and external conditions. This analysis of team performance is broadly similar to the analysis proposed for group performance by others many years ago (Steiner,1972) but there performance was analysed in terms of task demands, resources and the process used to achieve goals. Many analyses of this type have examined performance in a number of settings and found that team or group performance is only as good as the performance of the best individual performer or an average of group performance. However, Hartley (1997) has suggested that given the right conditions performance increments can be seen in groups when the process measured as the quality of communication is more effective in both simple and complex tasks.

In engagements team members may be lost, the task characteristics may not be those originally planned for, and the external conditions may change as the plan evolves. Communication helps ensure that resources and demand are adequately appraised prior to action but it can directly support the seven categories of team function identified by Fleishman and Zaccoro (1992) during action.

Fleishman and Zaccoro's (1992) functions of communication is broadly similar to the simple description of roles for communication put forward by Kanki and Palmer (1993). Kanki and Palmer (1993) suggested that communication provides information, establishes relationships, establishes predictable behaviour patterns, maintains attention to the task and is a management tool. Although not a perfect match the functions of communication identified by Leplat (1991) are a close match to four of the roles proposed by Fleishman and Zaccoro (1992). What is required is an empirical study to determine the types of communication actually used in the cockpit and their frequency during different phases of flight or engagement.

Effective communication may be useful in preventing cognitive lockup which occurs when operators fail to revise situation assessment (De Keyser and Woods, 1990; Woods, Johannesen, Cook and Sarter, 1994) but it could help in the overall decision making process. Communication can fulfil the roles identified by Jensen (1995) as important to effective communication in cockpit/crew resource management and support more effective decision making.

- Inquiry
- Advocacy
- Conflict Resolution
- Critique

It is important to be aware that the contribution of communication in normal operations may be invisible. It is only by removing the free availability for communication and challenging the participants with novel situations that one might reveal the true value of communication. Different types of experimental manipulation may reveal the importance of different types of communication with coordination functions important in some circumstances and orientation functions in others.

Some of the functions suggested for communication and their likely effects on decision making have also been suggested before by workers like Silverman (1992) in discussions of Human-Computer critiquing systems. Silverman (1992) suggested three methods of operation for decision support systems which may equally be applied to human-human decision support. The three methods were influencing, debiasing and directing. Director assistants seem to share much in common with the original Airbus philosophy of automated support in that the system and its moding were always right. Directing assistants assumes the existence of a very complete model of the task to be achieved and the capability of the resources to achieve it. Influencing and de-biasing assistants, are low level and intermediate strength critics which attempt to steer operators away from ineffective action or to present a case for a specific course of action. While decision support systems are approaching the latter level of competence in current technology they still cannot afford the contextual sensitivity of another experienced human decision maker in the same situation. In simple terms automated support can never use the same richness of information and provide the quality of advice that another experienced pilot can.

Some of the resources available to the military team such as combat system capability and aircraft manouevrability are well known and well understood prior to engagement but as time passes the resources can be depleted or suffer damage changing their capability. Thus, re-appraisal of the situation must take place and this requires communication of the current status. In terms of resources it is clear that automatic communication between participants revealing type and number of weapons' systems can be revealed automatically, by intelligent agents supported by on-board systems. Other parameters that might limit the operational role of aircraft such as fuel status could be automatically made available to other participants who might seek support in further engagements. Systems status could be made available automatically enabling participants to be made aware of limitations of other aircraft which might affect operational effectiveness e.g. decoy or defensive aids system status.

Thus, one goal of communication might be to create a big picture of the local environment, the status of the participants and the way in which the mission related activities are to fit within that framework. This ambitious goal may be over-optimistic in truly distributed decision making environments because an accurate global picture may be impossible to maintain when the number of independent active agents is beyond a critical level. Thus, a more plausible role is the provision of information between participants that allows them to decide individually how to manage the situation and use the available resources to greatest effect. This less ambitious goal is in accord with the view that shared knowledge is impossible because several agents in such a distributed system knowing the same facts need not interpret them in the same way. All that is required is a knowledge of others' intent in order to pragmatically align activities and act as if there was a shared consensus.

Adopting the more modest set of goals for communication it may be possible to envisage a way in which independent operators may communicate with each other and use on-board automation to improve their effective performance. By entering their intended plans or sequences of actions into a computer the operator can have their future plan reviewed and at the same time transmitted to

their associates on other platforms. All operators would effectively carry out similar operations and any conflicts in their individual plans could be revealed in the synthesized solutions produced by inter-agent communication. It is not envisaged that this method of operation is optimal but it may meet certain operational demands.

It reduces the need for two independent processes.

It would allow for repair in the event of an agent's removal.

It makes the information available on demand.

It may reduce the need for other forms of communication.

Failures in Communication

It is important to recognize that communication can fail to produce the effects that one expects because participants in high workload and high stress team performance situations frequently share information that is of low value. Information shared is frequently already known to the participants and there is some evidence to suggest that information that is ambiguous in form, interpretation or value is resource intensive to distribute as it needs further qualification. Sharing of well understood information does not encourage reviewing of the current planned response or increase team situational awareness in a way that prevents biases from dominating decision making. Indeed, communication which re-asserts known information may simply support development of a confirmation bias across a team in the form of Groupthink originally coined by Janis (1982).

Role of Communication in Critiquing

Silverman (1992) suggested a number of ways in which automated agent-based systems and decision support systems could collaboratively assist human operators and prevent them from making typical errors.

First, by asking leading questions a person may seek to influence the judgement of another by directing attention towards specific aspects of the environment or making the decision maker aware of the alternatives. Second, by repeating information one individual may influence a decision maker by inquisitively by directing the information, the information source or the interpretation towards further scrutiny. Or, the supporting individual may actively seek the other decision maker to adopt another action, another interpretation of the evidence or query the underlying reasoning.

Workload and Communication

One of the most serious problems for team activities is the effect of workload and recent experimental evidence working on aircrew under different levels of workload confirms that communication increases the cognitive burden during multi-task situations. Communication is a very demand intensive task and one which is frequently the first to suffer in the face of increasing workload. Communication can degrade in a number of ways because it relies upon coordination of transmission at a number of levels (Malone and Crowston, 1990). It can also fail to achieve its intended aim because it is received but not processed sufficiently. Communication can be lost because of inattention. and can be misinterpreted. Misinterpretation, limited processing and inattention seem broadly similar but they may have different effects on the sender and receiver in each case. Distortion, misunderstandings and repeated transmissions may result and these may interfere with other team activities. This degradation of team process behaviours associated with high workload has be accepted by senior researchers in the applied domain (Bowers et al., 1996).

Conclusions

It is clear that much remains to be explored with regard to the role of communication in military teams. In recent years there have been conflicting reports of high frequency communication as an index of effective team-performance and low frequencies of communication as an index of an effective shared mental-model. There are very few studies which have examined the frequency and nature of the communication in actual practitioners to determine the role of communication in effective team performance. Even fewer have tried to experimentally examine key issues relating to the effects of new systems or technologies even though some have raised doubts about their possible effects. It is surprising that aerospace designers seem reluctant to learn from the lessons derived from the application of automation in terrestrial and surface distributed systems which have provided many fine examples of the failures in cooperative and collaborative systems design.

There is no more fitting way to conclude this discussion of importance of communication requirements analysis in systems development than by quoting Schmidt (1991) on the deceptive nature of cooperative action in computer supported collaborative systems.

"The innocence and familiarity of cooperative work is deceptive. Cooperative work is difficult to bridle and coerce into a dependable model."

Seven years later the problems with the analysis of cooperative working are no less tractable. For this reason system developers and technologists developing solutions for environments with a communication requirement must support careful empirical evaluation of new interface and support technologies to ensure that the problems prevalent during the introduction of the Glass Cockpit are not re-created anew. It is not enough to adopt user-centred design because in multi-user systems the team must adopt the system or it will be doomed to failure.

References

Barrie D. (1998) Class War. Flight International, 30-31, 25-31st March.

Beach L.R. (1990) Image Theory: Decision Making in Personal and Organisational Contexts. Chichester : Wiley.

Beach L.R. and Mitchell (1990) Image Theory : A behavioural theory of decision making in realistic settings. *Organizational Behaviour and Human Performance*, 14, 10-59.

Beach L.R. (1991) Broadening the definition of decision making: The role of prechoice screening of options. *Psychological Science*, 4(4), 215-220.

Boehm-Davis D.A. (1990) Knowledge Elecitation and Representation. In J.I. Elkind, S.K. Card, J. Hochberg and B.M. Huey (Eds.) *Human Performance Models for Computer-Aided Engineering.* London : Academic Press.

Bowers C.A., Oser R.L. Salas E. and Cannon-Bowers J.A. (1996) Team Performance in Automated Systems. In R. Parasuraman and M. Mouloua (Eds.), *Automation and Human Performance : Theory and Applications.* New Jersey : Lawrence Erlbaum Associates Publishers.

Brehmer B. (1991) Some notes on the literature. In J. Rasmussen, B. Brehmer and J. Leplat (Eds.) Distributed Decision Making : Cognitive Models for Cooperative Work. London : John Wiley and Sons.

Cranmer C. and Cook M.J. (1997) Speech interuption effects in multi-modal tasks. Prospects for Spoken Language Technology. C5 (Human-Computer Interaction) Digest No.: 97/138.

Cook M.J., Elder L. and Ward G. (1997a) Decision Making, Planning and Teams. In C5 (Human-Computer Interaction) Digest No.: 97/137.

Cook M.J., Elder L. and Ward G. (1997b) Teamwork Aspects of Joint Tactical Decision Making. In C5 (Human-Computer Interaction) Digest No.: 97/137.

Cook M.J., Cranmer C., Milton C.A., Finan R.F. and Sapeluk A. (1997c) Hidden Usability Issues in Speech Interfaces, in AGARD Conference Proceedings 596, Audio Effectiveness in Aviation, AGARD-CP-596.

Cook M.J. (1997) Quick Thinking and Decision Making, In *Decision Making and Problem Solving*, Digest No. 97/366. London : IEE.

De Keyser V., and Woods D.D. (1990) Fixation errors : Failures to revise situation assessment in dynamic and risky systems. In A.G. Colombo and A. Saiz de Bustamante Eds.. System Reliability Assessment. Kluwer Academic : Netherlands.

De Sanctis and Gallupe (1987) A foundation for the study of group decision support systems. Management Science, 33 (5), 589-609.

Donnelly D. (1997) Decision Making on the Flight Deck. In Proceedings of IEE Colloquium on Decision Making and Problem Solving. Digest 97/366. London : Institute of Electrical Engineers.

Donnelly D. (1998) Decision Support on Civil Aircraft Flight Decks. Unpublished Masters Thesis. University of Bristol.

Eiser J.R. (1986) Social Psychology : Attitudes, cognition and social behaviour. Cambridge : Cambridge University Press.

Gasser L. (1991) Social conceptions of knowledge and action: Distributed Artificial Intelligence foundations and open systems semantics, *Artificial Intelligence*, 47, 107-138.

Gerson E.M. and Star S.L. (1986) Analyzing due processes in the workplace, *ACM Transactions on Office Information Systems*, 4(3), 257-270.

Grudin J. (1988) Why CSCW applications fail: problems in the design and evaluation of organisational interfaces. *CSCW Proceedings* 1988, 85-93 New York : ACM.

Hartley P. (1997) *Group Communication.* London : Routledge.

Isby D.C. (1997) *Jane's Air War 1 : Fighter Combat in the Jet Age.* London : Harper Collins Publishers.

Janis I. (1982) Victims of Groupthink : A psychological study of foreign policy decisions and fiascos 2nd Ed. Boston: Houghton and Mifflin.

Kanki and Palmer (1993) Communication and Crew Resource Management. In E.L. Wiener, B.G. Kanki and R.L. Helmreich, *Cockpit Resource Management.* London : Academic Press.

Kantowitz B.H. and Campbell J.L. (1996) Pilot Workload and Flightdeck Automation. In R. Parasuraman and M. Mouloua (Eds.), *Automation and Human Performance : Theory and Applications.* New Jersey : Lawrence Erlbaum Associates Publishers.

Leplat J. (1991) Organisation of Activity in Collective Tasks. In J. Rasmussen, B. Brehmer and J. Leplat (Eds.) Distributed Decision Making : Cognitive Models for Cooperative Work. London : John Wiley and Sons.

Malone and Crowston (1990) What is coordination theory and how can it help design cooperative work systems ? CSCW Proceedings 1990, 357-370. New York :ACM.

Mantovani G. (1996) *New Communication Environments : From Everyday to Virtual.* London: Taylor Francis.

Malin J.T. and Schreckenghost D.L. (1992) Making Intelligent Systems Team Players : Overview for Designers (NASA Technical Memorandum 104751). Houston, TX: Johnson Space Centre.

Malin J.T., Schreckenghost D.L., Woods D., Potter S., Johannesen L., Holloway M., and Forbus K. (1991) Making Intelligent Systems Team Players : Case Studies and Design Issues. Vol. 1: Human-Computer Interaction Design; Vol. 2.: Fault Management System Cases (NASA Technical Report 104738). Houston TX: Johnson Space Centre.

Mosier K.L. and Skitka L.J. (1996) Human Decision Makers and Automated Decision Aids : Made for Each Other ? In R. Parasuraman and M. Mouloua (Eds.), *Automation and Human Performance : Theory and Applications.* New Jersey : Lawrence Erlbaum Associates Publishers.

Parasuraman R., Mouloua, Molloy and Hillburn B. (1996) Monitoring of Automated Systems. In R. Parasuraman and M. Mouloua (Eds.), *Automation and Human Performance : Theory and Applications.* New Jersey : Lawrence Erlbaum Associates Publishers.

Prince C. and Salas E. (1993) Training and Research for Teamwork in Military Aircrew. In E.L. Wiener, B.G. Kanki and R.L. Helmreich, *Cockpit Resource Management.* London : Academic Press.

Scerbo M.W. (1996) Theoretical Perspectives on Adaptive Automation. In R. Parasuraman and M. Mouloua (Eds.), *Automation and Human Performance : Theory and Applications.* New Jersey : Lawrence Erlbaum Associates Publishers.

Schmidt K. (1991) Riding a tiger, or Computer Supported Cooperative Work. In Bannon L., Robinson M. and Schmidt K. (Eds) E-CSCW'91, Dordrecht: Kluwer.

Silverman B.G. (1992) Critiquing Human Error : A Knowledge Based Human-Computer Collaboration Approach. London: Academic Press.

Steiner I.D. (1972) Group Processes and Productivity. London : Academic Press.

Straus R. and Cooper R. (1989) Crew structure, automation, and communications. Proceedings of the Human Factors Society 33rd Annual Meeting (pp. 783-787). Santa Monica, CA: Human Factors Society.

Trevino L.K., Daft R.L. and Lengel R.H. (1990) Understanding managers' media choices: A symbolic interactionist perspective, In Fulk J. and Steinfield C. (Eds) Organisations and Communication Technology, pp. 71-94, Newbury Park, CA: Sage.

Wickens C.D., Marsh R., Raby M., Straus S., Cooper R., Hulin C.L. and Switzer F. (1989) Aircrew performance as a function of automation and crew composition: A simulator study. Proceedings of the Human Factors Society 33rd Annual Meeting (pp. 792-796). Santa Monica, CA: Human Factors Society.

Wickens C.D. and Huey B. (1993) Workload Transition : Implications for Individual and Team Performance. Washington D.C.: National Academy Press.

Woods D. (1990) Modeling and predicting human error. In J.I. Elkind, S.K. Card, J. Hochberg and B.M. Huey (Eds.) *Human Performance Models for Computer-Aided Engineering.* London : Academic Press.

Woods D.D., Johannesen L.J., Cook R.I. and Sarter N.B. (1994) Behind Human Error : Cognitive Systems, Computers, and Hindsight. CSERIAC, Wright Patterson Air Force Base : Ohio, U.S.A..

Form of Communication	Standardized Format	Frequency of Communication	Usage
Simple - action based coordination.	Agreed standard or unit level standard - low in precision.	High	Coordination of moment to moment actions.
Plan Based - role based coordination.	Free form	Low	Coordination, maintenance and substitution of future actions.

Table 1 : Levels of communication.

Communication Acts	Processes and Resources Used
Receipt and interpretation of communication. *Provide information when asked.* *Repeat information.* *Convey information concisely.* *Reply without a question or comment.*	Diversion of attention to incoming information. Analysis of information. Consideration of response.
Validation speech acts confirming receipt. *Acknowledge communication.* *Repeat information.* *Make no response.*	Generation of acknowledgement and receipt of confirmation of acknowledgement.
Generation of speech and management of delivery *Use standard terminology.* *Convey information concisely.*	Preparation of information and selection of time for delivery. Negotiation of delivery. Confirmation of comprehension by others and /or receipt.

Table 2 : Communication acts, the processing required and resources used. Text in italics refers to communication task components from Prince and Salas's (1993) paper.

Function of Communication	Error Correction Type and Activity
Orientation - the collection and distribution of information. (See also Leplat, 1991).	Ensures availability of knowledge. Effective problem representation. To prevent representation errors associated with incomplete or incorrect mental models. To prevent confirmation bias.
Resource distribution - matching of demand to available resource.	To prevent overloading of individual operators. Preventing mistakes in choosing between alternatives.
Timing - monitoring and coordination of the pace of activities. (See also Leplat, 1991).	To prevent missing side-effects like collisions or failed interceptions.
Response coordination - sequencing and timing of individual contributions. (See also Leplat, 1991).	To prevent missing side-effects like collisions or failed interceptions.
Systems monitoring - management of progress at team and individual levels. (See also Leplat, 1991).	Articulation of both team and individual goals to prevent a narrow focus in operators under heavy demand.
Motivation - communication of team objectives and performance norms.	To prevent thematic vagabonding under high demand and time pressure.
Procedural maintenance	To prevent lapses or slips of action, capture - substitution errors, omission errors and mode errors.

Table 3 : Functions of communication adapted from Fleishman and Zaccoro (1992) and common errors adapted from Woods (1990) which may be reduced by communication.

PRACTICAL RESULTS OF THE VERIFICATION OF THE MODEL OF A MULTIOPERATOR SYSTEM

Sýkora J., Bahbouh R., Øadová J., Dvoøák J., Podivinský I., Chamrád P.
STRESS RESEARCH CENTER
Kbely Airport
19706 Praha 9, Mladoboleslavska
Czech Republic
Tel: 02/20 20 74 44, 02/20 20 74 57, 02/20 20 74 40
Fax: 02/850 15 32
--
High Military College of the Ground Forces
of the Ministry of Defense
Vyškov, Czech Republic

The Stress Research Center is a specialized laboratory aimed at analyses of individuals and small groups of humans under stress, mainly under conditions of extremal situations in the Czech Army Ground Forces and in the Czech Air Force.

The results presented here are based on experiments performed in the past ten years.

The understanding of human/human and human/system interface issues team communication, aided by a suitable model, could increase the reliability of the crew and multicrew system operations. As yet, thus was rather difficult on account of insufficient possibility of quantification of the relations of subjects, members of the system.

A quantified, dynamic model of psycho-social, intra- and intergroup relations was realized by methods of dynamic sociometry, based on fuzzy set and image theory, developed by Bahbouh in our Laboratory. The model was verified in practice in the E.S.A. HUBES-94 experiment and in the Czech Air Force.

METHODS AND RESULTS:

Methods used are non-invasive, such as color assignment test, questionnaires and video- and audiotaped behavior and communication analyses.

Our work is based on biological, psychological and sociological approach, summarized under the concept of the BIOPSYS philosophy.

Following methods were used in the HUBES-94 experiment:

A) Pencil-paper measures:
- Personal Views Survey
- Test of Actual Harmonization Degree
- Test of Color Assignment

B) Physiological parameters analysis:
- Physical activity
- Skin conductance reaction
- Heart rate and its variability

C) Video- and audiotape recordings analysis
- Number of contacts and subjective preferences
- Quality of communication, formal and informal

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

The model is a quantitative expression of positions and relations of subjects (humans or groups) in a complex system, leading in its graphical form to what is called a social map. Altitudes (in different colors as used in geographical maps) represent subjects' social positions (high or low), distances are social distances, isohypses represent the levels of social relations.

EXPERIMENTAL VERIFICATION:

The work is based on results of experiments with groups of volunteers (three to fifteen) under stress of one to twenty two weeks of social deprivation, simulating space orbital mission conditions. One of the experiments was realized as participation in the E.S.A. HUBES-94 program. Subsequently the model was verified in two studies of social relations, in one test-pilot squadron and among all units of one Czech Air Force corp. The aim of this part of the work was testing of the quality of the flow of commands and of command-units relations on all levels of the system under analysis.

As one of the results we formulated the concept of a twofold communication channel:

(1) The main communication channel, providing the flow of formal information, such as commands, orders, instructions etc. together with a corresponding feed-back communication - reports, administrative messages etc. This is evidently aimed at the organization and administration of the activities of the social system under consideration.

(2) The lateral communication channel, presenting non-formal information about the attitudes, motivations, feelings and moods of the members of the operational crew and groups systems. This is aimed rather at the moral aspects of the members of the social system analyzed.

Allow me to present some results of our work: from one TEST-PILOT GROUP, from the AIR FORCE GROUP and from the HUBES-94 group dynamics.

This is the social map of our TPG (Fig-a). The social relations within this group are presented in just the some way used in geographical (topographical) maps. Seven subjects formed the group. The altitudes represent the levels of social position of each subject as you see No. 5 is characterized by the highest altitude and No. 3 by the lowest one. This is in connection with their social roles. The distances on the map correspond to the social distances in the group. Differently from Moreno´s sociometry we are able to express asymmetric relations. This is viewed as a valleys on the map. Social map is the model of social group. It was verified after five years. The subject No. 5 is now studying in the Military University in USA, No. 3 (the Chief of staff) left the Army past year.

Next case (Fig-b) is the sociomap of the whole Czech Air Force in the situation of the year 93. (I apologize for presenting data from the time 5 years ago, as present data are declared as not open).

I hope, now you understand the principals of the social mapping.

The whole Air Force forms a relatively homogenous complex with 3 exceptions, the first one is KBELY, which can be understood, this is the transport aviation base and those a fighter and fighter-bomber units. CASLAV is a newly formed fighter and fighter-bomber base. The third is the case of the High Command and staff of the Air Force. Evidently the flow of information between the High Command

and AF units is profoundly disturbed. Please, pay attention to the fact that social relations between the Air Bases and High Command expressing sympathies and antipathies were rather catastrophic. Clearly the commander of this AF units complex the 3.corp was respected in the highest social position. No wonder that the commander of this complex is the Chief of the Air Force at present.

The question of the disturbed flow of communication were analyzed in much detail and proposals for social interventions were formulated and are in the process of realization now.

The last example is the dynamic of social relations in an experimental group of three subjects under conditions of a simulated 153 days space mission.

(FIG 1) Our subject formed a homogeneous group with a high level of mutual social position before the experiment.

(FIG 1-7) This profoundly changed in relation with critical situations during the course of the experiment.

CONCLUSIONS:

A set of social models taken at different time intervals gives a dynamic insight into the development of social relations in the system under study. Their quantitative evaluation and prediction of future development are thus possible.

The structure of social mapping
The real situation of Stress Squadron

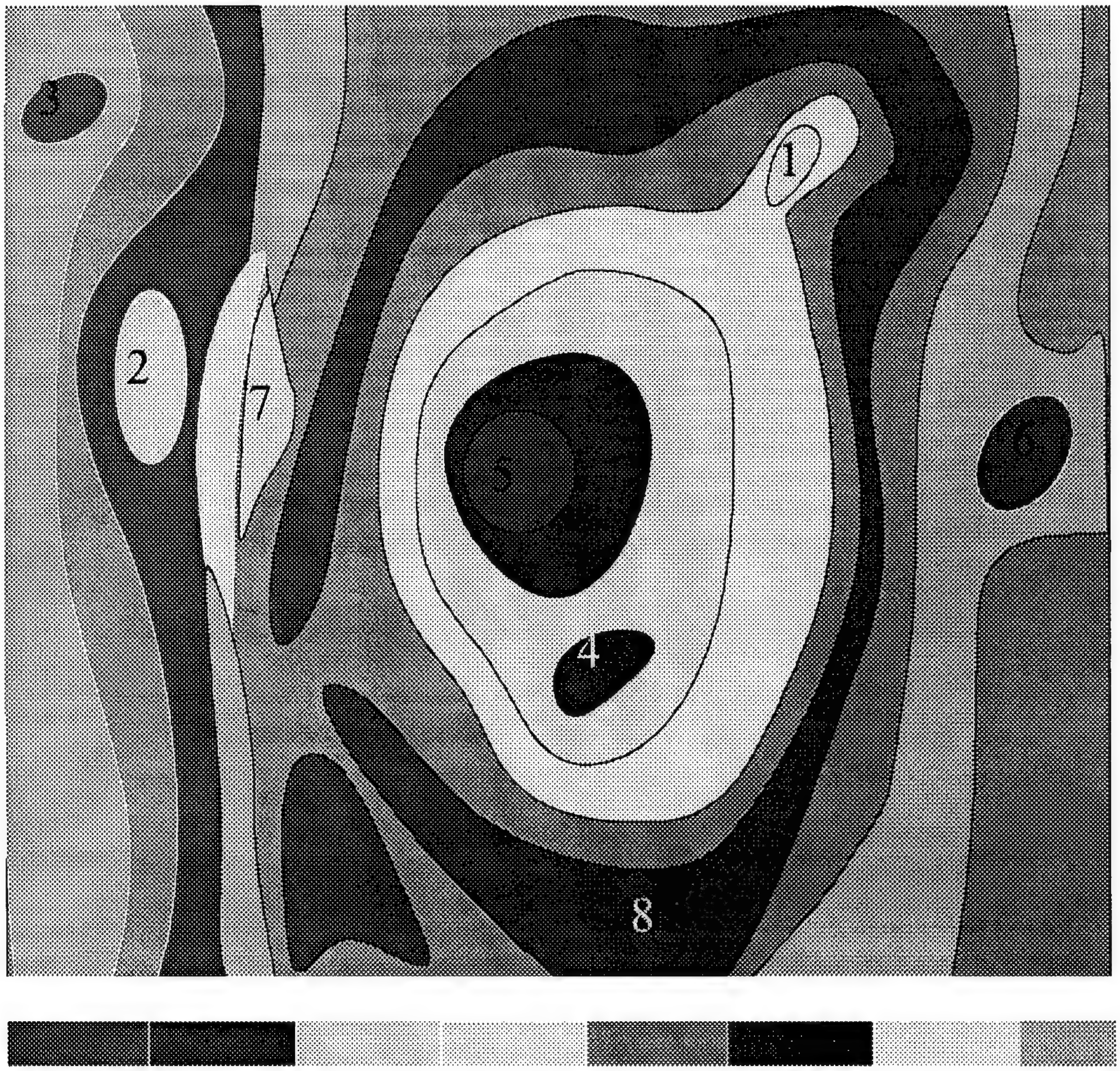


Fig. A

Sociomapping in Czech Air Force

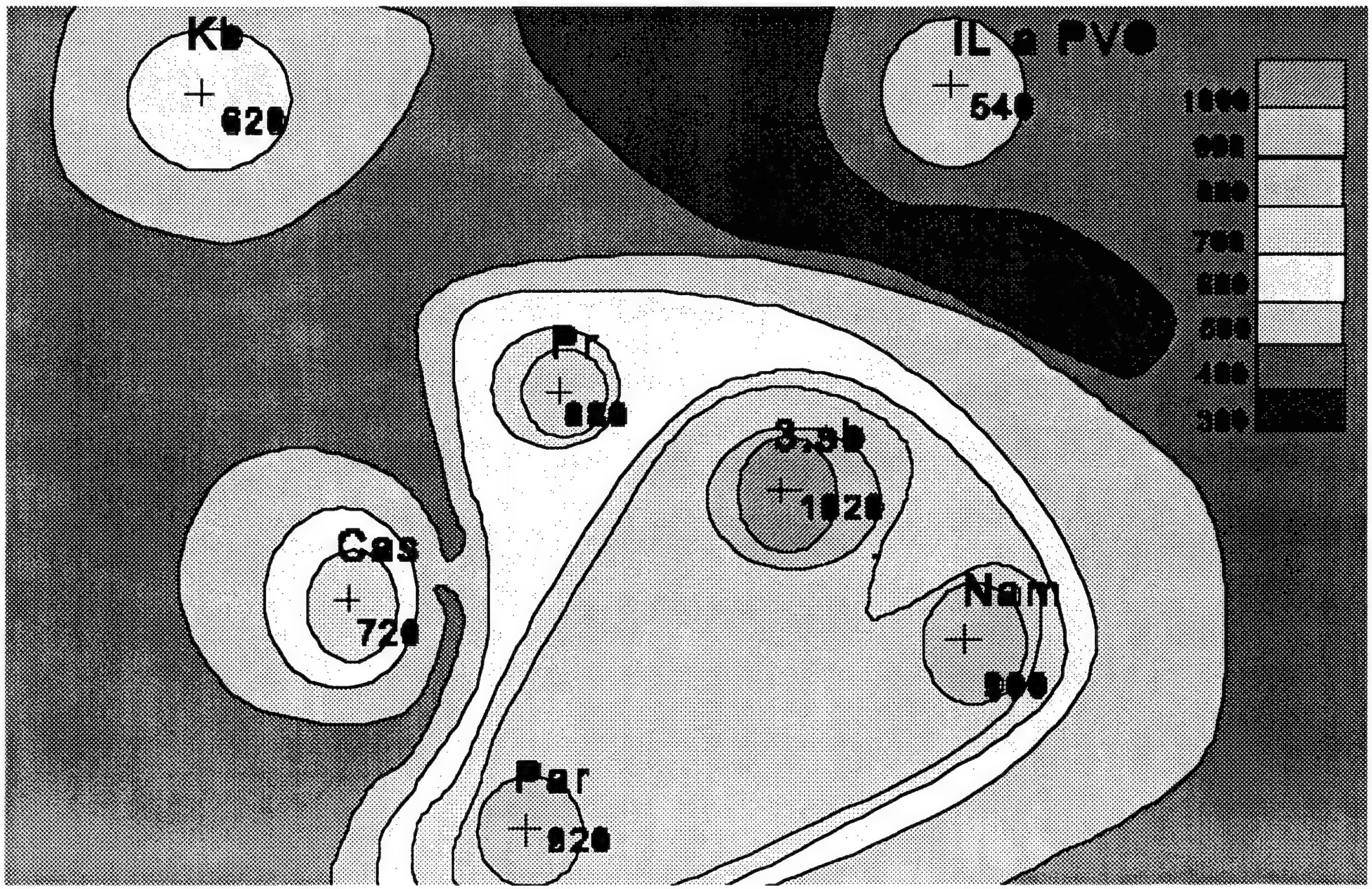
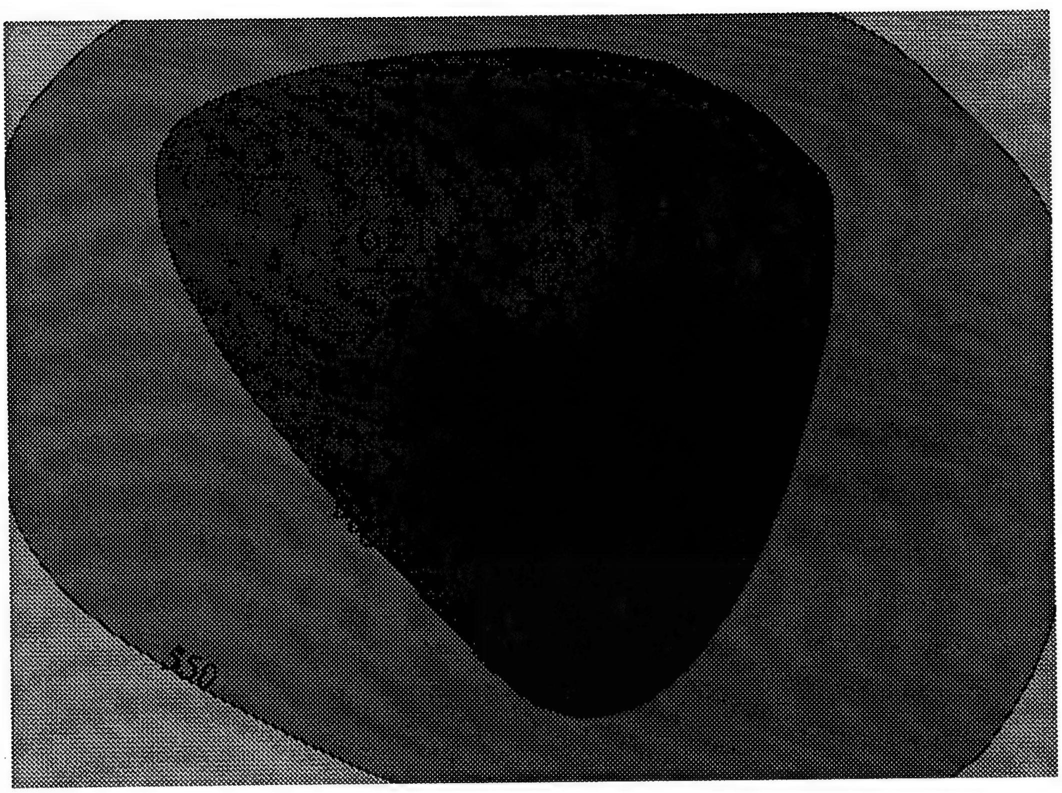


Fig. B

Fig. 1

28.7.1994



THE CHANGES OF INTERRELATIONS IN A GROUP DURING LONG TERM ISOLATION

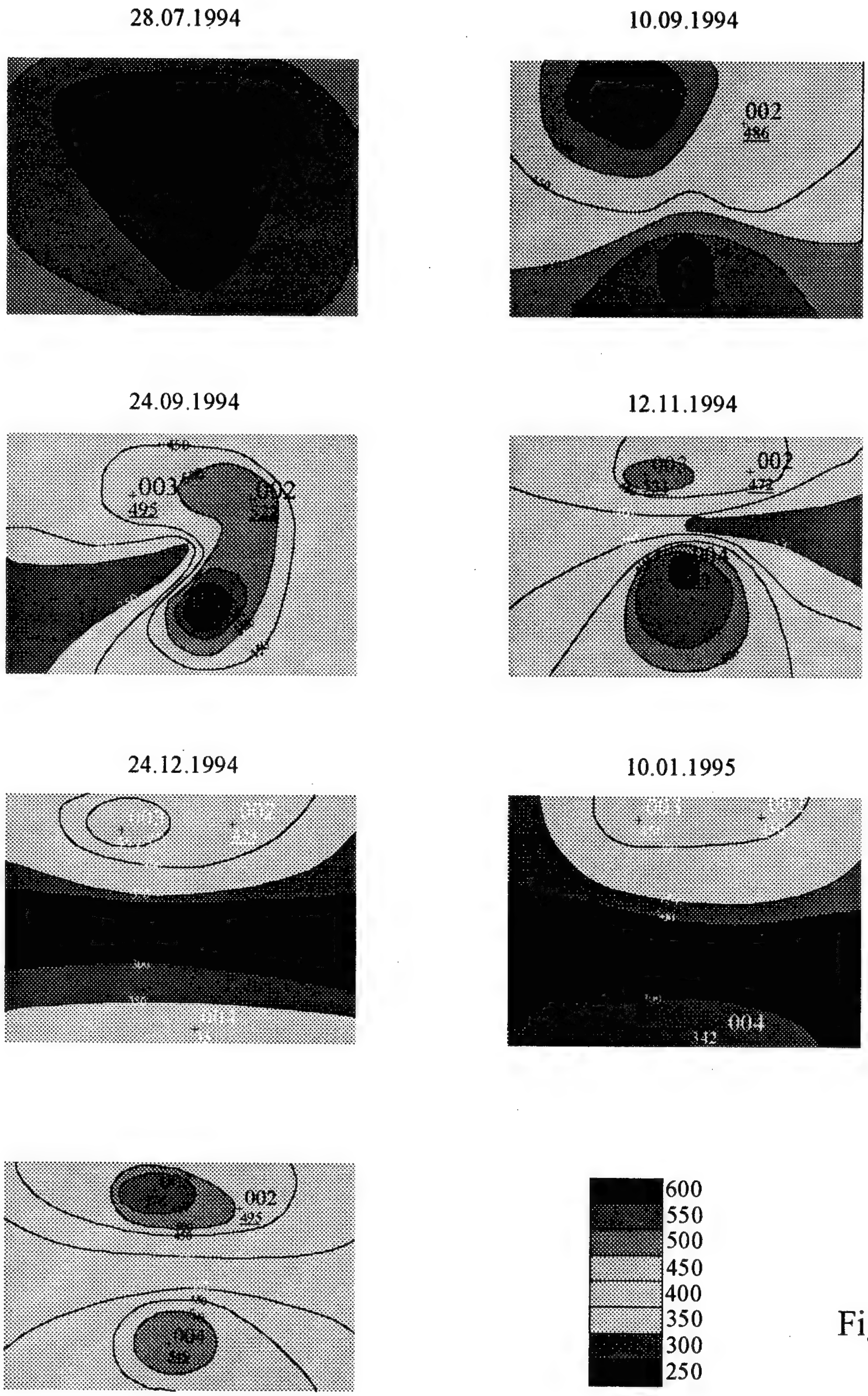


Fig. 1-7

Social Organization of Knowledge in Teams: Issues for Computer Support

John T. Nosek
Computer & Information Sciences Dept. (38-24)
Wachman Hall, Rm 303
Temple University
Philadelphia, PA 19122-6094
USA

1. SUMMARY
Within many domains, complexity encompasses many nuances of ill-definition, fluidity, organizational variation, uncertainty, conflicting constraints, and multiple solutions. Responses to these areas of complexity necessitate the social construction of knowledge among various multidisciplinary team members. Multi-theoretical foundations for group sensemaking are explored. Group Cognition is offered as the basis of all cognition and is explained as a combination of Distributed and Coordinated Cognition that directly affects the creation/recreation of distributed and similar knowledges within a team. Based on these foundations, initial guidance is offered for augmenting the social construction of knowledge and artifacts. It is important to remember that this report must be considered a work-in-progress, a "snapshot" of one exploration of a very complex subject.

2. INTRODUCTION
Within many domains, complexity encompasses many nuances of ill-definition, fluidity, organizational variation, uncertainty, conflicting constraints, and multiple solutions. Responses to these areas of complexity necessitate the social construction of knowledge among various multidisciplinary team members. The process by which interpretation, meaning, decisions, and actions transpire is referred to as *group sensemaking*. This process is especially salient for understanding how to achieve a group-centered approach in the design of multi-person/team interfaces. Many of the current information technologies have not been designed from a group-centered approach, which limit their usefulness.

This paper focuses on the multi-theoretical bases for group sensemaking in the social construction of knowledge. Group Cognition is offered as the basis of all cognition and is explained as a combination of Distributed and Coordinated Cognition that directly affects the creation/recreation of distributed and similar knowledges within a team. Based on these foundations, initial guidance is offered for computerized support of the social organization of knowledge within teams.

3. FOUNDATIONS
Sensemaking is "the process whereby people interpret their world to produce the sense that shared meanings exist [Leiter in Gephart, 1993, pp. 1469-1470]." Social actors actively engage in sensemaking by interpreting the social world through conversation and textual accounts, explanations offered and accepted, and ongoing discourses that describe and make sense of the social world [Gephart, 1993, Weick,1979]. Sensemaking occurs and can be studied in the discourses of social members – the intersubjective social world – rather than simply occurring in their minds. Further, the socially constructed object, or facts, of the world exist through and are located in the discursive sensemaking of members [Gephart, 1993, 1470]."

In complex environments, where not all variables and relationships are known, humans create rather than discover their future [See Figure 1]. They create the future by accepting stimuli from their environment, including others around them, and interpreting what these stimuli mean. The subsequent actions, including probing of the environment, leads to additional stimuli that must provide meaningful affordances to grab attention, and subsequent processing. Human and non-human agents must be attuned to relevant affordances, to interpret them, to act based on them, and to probe for additional stimuli.

3.1 Constructionist and Ecological Perspectives
Both ecological and constructionist views offer useful aspects that help explain how actors interpret their environment [Preece et al , 1995]. Those who argue for the ecological view emphasize that observable objects afford their meanings in actors without conscious interpretation. Constructionists argue that actors observe stimuli and construct their meaning.

Figure 2 below provides a synthesis of these viewpoints and introduces some qualifications of terms to support this synthesis. One may argue that looking at the characteristics of the object alone within its background, one may say that the object projects an affordance, a "projected affordance." In Figure 2, Stimulus A and Stimulus B possess characteristics that project their meanings. Stimulus A projects a weaker affordance, while Stimulus B projects a stronger affordance. The characteristics of the observer affect what affordances of the object are received, "received affordances." Focussing on Stimulus B, the actor on the left possesses characteristics that enhances the projected affordance, causing a stronger "received affordance," while the actor on the right possesses characteristics that diminishes the projected affordance, causing a weaker "received affordance." Depending on the situation, the received affordance can then be interpreted and a meaning constructed from the received affordance, "interpreted affordance." The actor on the left interprets the received affordance, further enhancing its meaning, while the actor on the right interprets the received affordance causing a diminished meaning. For example, assume Stimulus A is a stick while stimulus B is a standard doorknob. Assume the stick projects a weaker signal

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

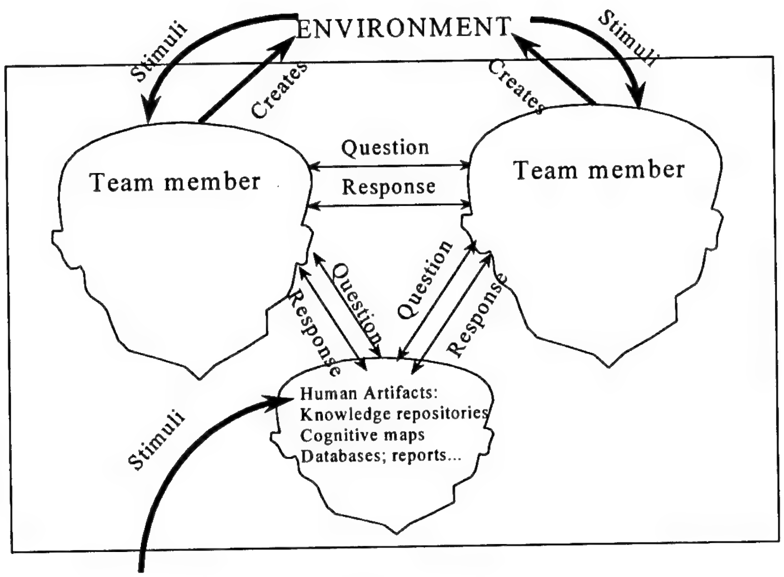


Figure 1
Group Sensemaking in Ill-defined Situations

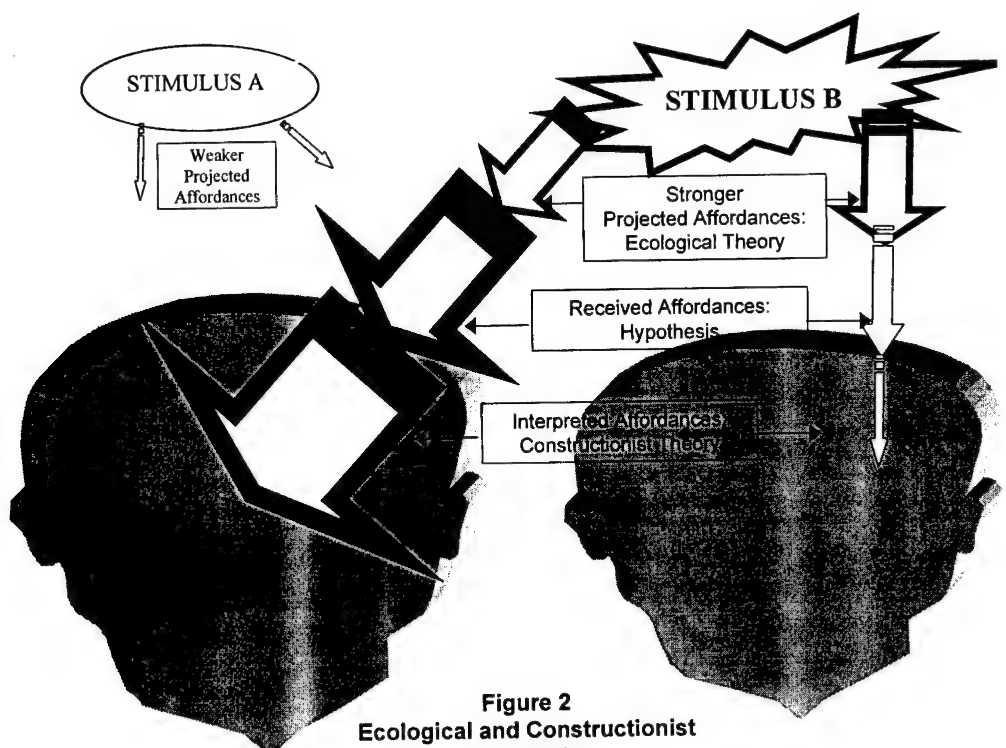


Figure 2
Ecological and Constructionist Synthesis

of what to do with it, while the doorknob by its design affords that the object is to be grabbed. The doorknob projects a certain affordance regardless of the characteristics of the observer. However, let's assume the actor on the left has keen eyesight, while the actor on the right does not see well or is blind, then the "received affordance" is enhanced for the actor on the left while it is diminished for the actor on the right. The projected affordance of Stimulus B remains the same, but the characteristics of the actors affect the resulting affordance received by them. The actor on the left interprets the clear affordance of the object, perhaps compares it with previous situations, further enhancing the received affordance, the "interpreted affordance," and easily grasps the doorknob. The actor on the right can't see the doorknob or distinguish it from the surface around it so the received affordance is further dimished, and the "interpreted affordance" results in little or no thought to grab the doorknob. It seems reasonable to view resultant meanings of object as relative to the object within its surroundings and relative to the characteristics of the observer within his/her/its task environment.

Extending these notions to the subject of the report, it seems reasonable to accept the ecological viewpoint when the projected affordance is strong, the situation is less ambiguous, and the actor does not possess characteristics that would prevent the reception of affordances. At the same time, it appears reasonable that when the situation is more ambiguous and the signal is weaker, that more interpretation is required and the constructionist viewpoint predominates.

Almost by definition, in ambiguous situations with high equivocality, the projected and received affordances must be weak. The situation is not well defined enough for the reception of signals in a non-ambiguous way. The actor/s must create a fiction in order to make sense of the world sufficiently. They hope this fiction created provides for a reasonable future. In these situations actors who perceive clear meaning and strong signals from stimuli are most likely being affected by cognitive and social biases. This could be a more dangerous situation and reinforce erroneous directions. In these situations it is more prudent to accept the notion of weak projected affordances that require cautious collaborative interpretations and vigilant scanning of the environment to interpret reactions to actions taken by relevant actors.

To summarize, stimuli meanings are relevant to the actors and the observable stimuli within a given situation. Ecologists and Constructionists both contribute to better understanding how meanings are ascribed to objects in a given situation.

3.2 Boundary Objects

With greater shared context (shared beliefs, expectations, perceptions...) intent is casually communicated. Perspective Taking occurs through boundary objects [Boland and Tenkasi, 1995]. Boundary objects are "anything perceptible by one or more of the senses [American Heritage Dictionary, 1980]," i.e., anything that can be observed consciously or subconsciously. Ethnographers view sensemaking dialogs as a way to externalize thoughts and achieve a shared construction of meaning. These dialogs may be considered boundary objects that permit exchange of thoughts. Mapping techniques, as described earlier, can also be considered boundary objects.

Non-verbal expressions can be classified as boundary objects. These include such things as "body language," tone of voice, raised heartbeats, head movement, eye movement, gestures, brain patterns, etc. Boundary objects can be used to identify convergence and divergence related to a given situation. For example, in highly dynamic situations, early signals of divergence may emerge at a subconscious level, but may be of insufficient strength or inchoate form to surface at a verbal level. Real-time monitoring of non-verbal boundary objects may provide early warning of divergence. For example, real time monitoring of speech patterns, eye, or head movements for members within a team may indicate early warnings of divergence and may be used to direct attention at a verbal level to consciously direct attention. The ease of capturing and screening boundary objects relates to the task characteristics.

3.3 Data, Information, and Knowledge

Knowledge is not about giving or getting [Senge, 1997]. Knowledge sharing is creating new potential/capacities for action [Churchman, 1971]. There is no difference in the physical nature of data and information. Informational value is relative to the capacity of an actor, human or non-human. The actor possesses the capacity to interpret the data so they have informational value for the actor. When data are interpreted as having informational value, they are labeled as information. As noted above, knowledge is the capacity to act, which includes conceptualizing. Therefore new knowledge is the increased capacity to act. This increased capacity to act may be situational; that is, the data that are interpreted to have informational value may provide an increased capacity to act for a given situation, but may not necessarily have increased the actor's capacity to act after the situation is over. On the other hand, actors that can learn may be able to increase their capacity to act in the future from these interpretive acts and may gain new knowledge that remains after the situational interpretation. For example, an airplane mechanic may be responsible for replacing a certain part when it shows problems. If this is a repetitive act and the mechanic is permitted to work on only one kind of problem, the mechanic may become automatic and his/her speed to perform the action may increase from this repetition. This would be an increased capacity to act; however, there could also be a point of declining performance if the repetition continued. Repeating the stimulus-response events may cause less attention to be paid to received affordances and the capacity to act may actually decline.

When one refers to information management, one means that data are organized in a manner to provide informational value to an actor who possesses the capability to act. Knowledge management is a broad term, however, at its crux, knowledge management has as its goal a way to reduce the energy required to interpret data in novel but similar situations. Usually, each time it is required, an actor, the external interpreter, interprets data that provides informational value for the actor to act. To transition from data to information, an external actor provides the interpretation of data to determine informational value and achieves a higher capacity to act, knowledge. This means the external actor provides the situational assessment and interpretation. To transition from data to knowledge without an external actor, knowledge-based/management systems must adequately describe a situation and match a stored interpretation to correctly act in a new situation with new data.

3.4 Individual responses to stimuli (Figure 3)

Figure 3, presents a diagram of an interpretation process. As discussed earlier, a stimulus projects a certain value of what it is, the projected affordance, the first arrow that emerges from the stimulus. Actors, with their own set of characteristics, subconsciously filter the projected affordances into received affordances, the second arrow emerging from the first arrow. Actors interpret the received affordances and create an "interpreted affordance", the third arrow emerging from the second arrow. The actor on the left interprets the received affordance, the interpreted affordance, as important, indicated by the larger arrow. This datum is interpreted as providing informational value, and this actor's knowledge, the capacity to act, increases. One could say, this actor has greater knowledge of the situation. The actor on the right interprets the received affordance, the interpreted affordance, as not important, indicated by the diminished arrow. This datum is interpreted as providing no informational value, and this actor's knowledge, the capacity to act, does not increase. One could say there are two "knowledges" of the same situation [Edamala, 1997], i.e., the actor on the left has increased his/her/its knowledge of the situation, and this knowledge of the situation is different from the knowledge that the actor on the right possesses. The skills, background, motivations... of actors affect interpreted affordance. People may selectively filter projected affordances and construct different meanings while converging on a similar way to act. Externally viewed, the convergence to act in a similar way may falsely indicate similar capacities to act (knowledge), or similar mental models, however, multiple knowledges or mental models are likely to exist. For example, two people cast a vote for someone, but their mental models could be different, even inconsistent [Shaw and Gaines, 1994] with each other, however they act in the same way, casting the same vote. This relates to how much mental models need to be shared or how similar the knowledges of the situation must be to act in a similar, coordinated way. Shaw & Gaines [1994] emphasize the importance of coordination over consistency in team action.

The more important the action, the more dynamic, equivocal the task, the more unreliable the data, the more important group sensemaking to the emergence of knowledges in this situation, the emergence of the capacities to act, sufficiently coordinated to engender effective action. In these cases, knowledges are likely different, but the emergence of sufficient capacities to act in a coordinated fashion is critically dependent on the social construction of these knowledges.

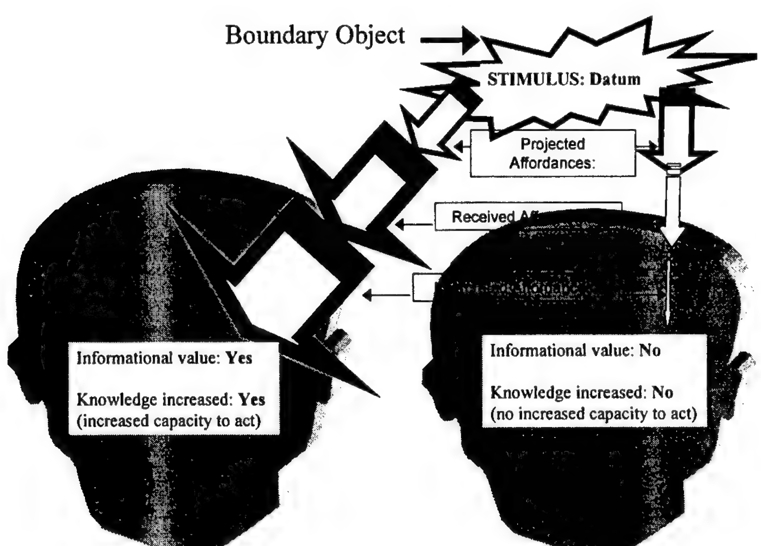


Figure 3
Individual Reponses to Stimulus

3.5 Group Cognition and Group Knowledge

Cognition is the process where capacities to act manifest themselves. Given a certain intelligence (capacity to acquire and apply knowledge), cognition is the mental process by which capacities to act are acquired and created. Over the last several years, my view of cognition has changed from one where all cognition is individual to where all cognition is group cognition. When I speak of all cognition being group cognition, it is a subset of the ideas of Winograd and Flores [1987] who state that all cognition is social and emphasize the role of language and society in one's thinking. At the most elemental level, individuals use words within their own minds and with others to think about something. This reflects the views of several researchers who emphasize that language strongly and directly affects thought. Language is a social artifact created and employed by a community of actors.

To distinguish from these broad ideas that all cognition is social, Group Cognition deals with the actual thoughts that are generated within one's mind. What we think depends upon one's interactions with the world, a world of other actors and actor-created artifacts. It is the boundary objects (anything observable) as initial stimuli, and the reflection on these objects, that stimulate the generation of thoughts, the cognition process. It is the cognition process that recreates the knowledge (the capacity to act) available in a situation. Therefore, "what one thinks" is dependent on these boundary objects that originate with actors, both human and non-human. Knowledge that has not been previously externalized and recorded, only exists at the moment of activation/recreation (tacit and explicit), i.e., tacit knowledge is a capacity to act that is activated, only one is not aware of it. Explicit knowledge is knowledge that is activated, and the actor is aware of it. One can externalize both tacit knowledge and explicit knowledge. Tacit knowledge emerges as observable capacity to act -- a by-product of actions/behaviors, including open reflection of these actions/behaviors.

From the Human-Computer Interaction Literature, Distributed Cognition has been used to describe the coordinated actions within a group [Preece et al, 1995]. Each team member has specialized roles and knowledge. There is some overlap of knowledge needed to achieve smooth coordination. Many times this overlap in knowledge is acquired and recreated tacitly through observation of boundary objects, which makes it difficult to discover. This definition of distributed cognition includes cognition and knowledges that need to be distributed (not replicated by individuals) and cognition that needs to be coordinated to effect similar knowledges. Therefore distributed cognition will be used here as including only cognition and knowledges that are distributed.

Cognition deals with the process of creating/recreating knowledge (capacity to act within a situation). Boundary objects and mental objects (tacit or explicit: one could be reflecting on objects subconsciously), affect cognition (the process to create/recreate knowledge (the capacity to act)) within working memory.

How do we coordinate cognition within a group to create reasonable knowledges of the situation? What boundary objects are needed and when and how do they need to be employed to create the knowledges of the situation to ensure effective action. Within a group there is a need for distributed knowledge and distributed cognition that are affected by boundary objects available to actors who receive projected affordances and interpret these affordances, tacitly and explicitly. There is also a need for more coordinated cognitive processing among group members who create similar knowledges of the situation. This is the essence of the social construction of knowledge and Figures 4a and 4b below depict this description. Figure 4a depicts one end of the continuum where group cognition and knowledge is completely distributed. At Time 1 in Situation A, the received affordances of stimuli are different as desired, the cognition process is different for the actor on the left than for the actor on the right. As a reminder, actors can be human and non-human. There are some characteristic differences in the actors that cause the projected affordance of the stimulus, Datum 1, to be received differently. For the actor on the left, the received affordance is enhanced, while for the actor on the right the received affordance is diminished. The cognitive process for the actor on the left is different then for the actor on the right, with the result that the interpreted affordance constructed by the actor on the left is enhanced while the interpreted affordance for the actor constructed by the actor on the right is diminished. For the actor on the left, the result is that datum 1 is interpreted as having informational value and he/she/it creates/recreates a capacity to act (knowledge).

For the actor on the right, the result is that datum 1 is interpreted as not having informational value and he/she/it does not create/recreate a capacity to act (knowledge). As noted previously, this is sometimes necessary and advantageous for a group to work in a distributed way. Each member performs specialized roles and needs specialized capacities to act. Figure 4b depicts the other end of the continuum, where group cognition is coordinated with the intent of creating/recreating similar knowledges of the situation for both actors. At Time 2 in Situation A, the received affordances of the stimuli are similar, as desired, as a result of the intent to create coordinated cognition processes in both actors. Characteristics of the actors, relevant to receiving the projected affordance of the stimulus, Datum 2, are sufficiently similar to cause the received affordance to be the same for both actors. In this case the received affordance is enhanced for both actors. The cognitive process for both actors are similar with the result that the interpreted affordance constructed by the actors is similar. For both actors the result is that datum 2 is interpreted as having similar informational value and the actor on the left creates/recreates a capacity to act (knowledge) similar to the actor on the right. By viewing the social construction of knowledge in this way, there are a number

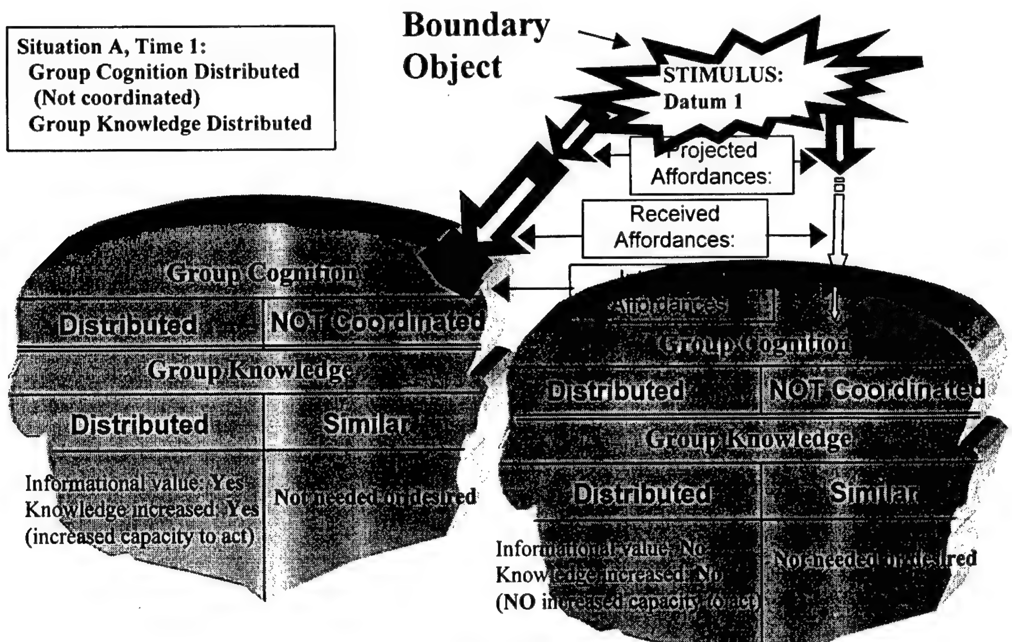


Figure 4a
Distributed Group Cognition & Knowleges

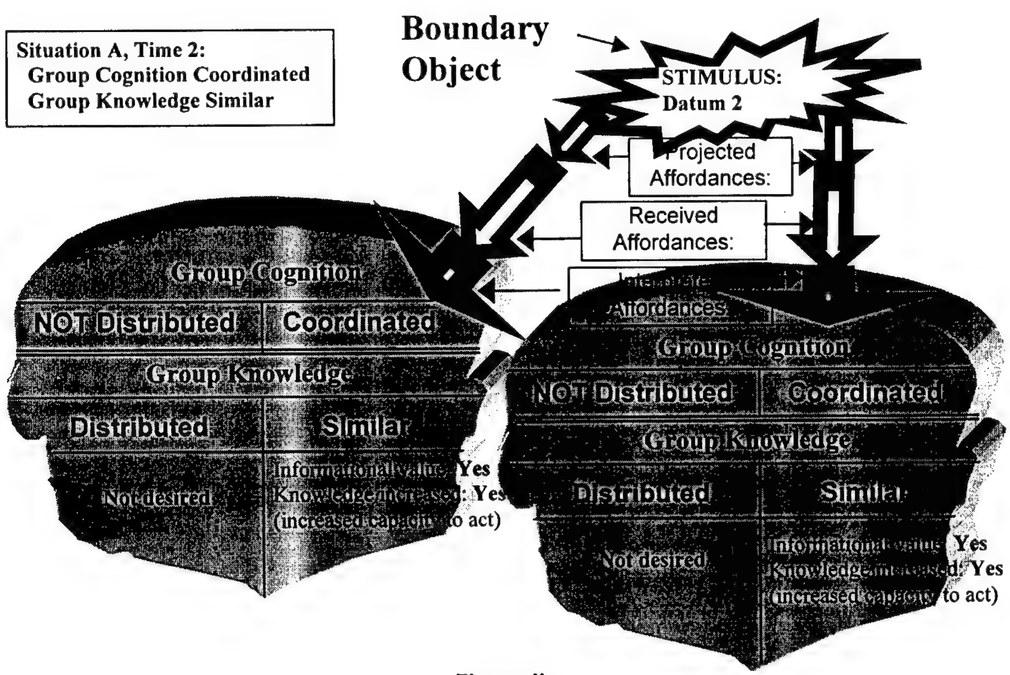


Figure 4b
Coordinated Group Cognition & Creating Similar Knowleges

of areas that can be addressed that will help us to better understand and support effective and efficient social construction of knowledges.

4. IMPLICATIONS AND EXTRAPOLATIONS FROM RESEARCH

The scope of this project prevented any thorough review of all existing technologies to support the social construction of knowledge and lower level technologies such as shared screens and pointing technologies have been reviewed extensively by other researchers [See CSCW literature for detailed studies]. Instead this section looks at what guidance we can glean from the previous section to better support the social construction of knowledge. A table format will be used for this section [See Table 3]. In some cases, where guidance appears to be contradictory an effort will be make to identify these. Finally, some guidance will warrant more complete discussion, and where possible this will be provided. As noted earlier, this report should be considered a work-in-progress and this section should be considered a starting point for a more definitive review.

Table 3
Implications for Supporting the Social Construction of Knowledge

Guidance	Bases/Comments
1. Capture rationales	Boland & Tenkasi [1995] Schmidt and Bannon [1992] Sage [1981] Nah and Kim [1997]
2. Capture stories/narratives	Boland & Tenkasi [1995]; Wynn [1997]
3. Reveal originator	Cyert and March, Strauss et al, Cicourel in Schmidt and Bannon [1992] Sproule [1980] **Contradicts 4:** to maintain anonymity
4. Maintain anonymity	Dennis et al **Contradicts 3:** to reveal originator
5. Provide contextual information	Schmidt and Bannon [1992]
6. Provide deniability	Schmidt and Bannon [1992] **Contradicts 1:** capture rationales; Supports 4: maintain anonymity
7. Provide information about originator without identifying originator	Nosek[1998, this report] **Compromise of 1:** capture rationales; **3:** reveal originator; **4:** maintain anonymity; **6:** provide deniability
8. Support reflection	Boland and Tenkasi [1995] Dreyfus & Drefus Shaw & Gaines [1994]
9. Treat social constructions, such as facts, as objects	Gephart [1993] Boland and Tenkasi [1995] Wynn [1997]
10. Identify increased divergence and convergence of attention to stimuli: • help identify affordances that may be stimulating this • provide means for groups to help create divergence and convergence • provide means to identify divergence within groups	Ackermann [1997] uses a conceptual mapping tool to gain convergence through consensus. She feels the best way to achieve this is through recognizing divergences and exploring the reasons for them. Decision Explorer (a concept mapping tool) is based on Kelly's Personal Construct Theory [1955]. Personal Construct Theory states that people have "opposite" or "opposing" limits to their views relative to some scale, e.g., when one says hot, the person implicitly and perhaps subconsciously has an idea of what "hot" is as compared to what "cold" is. One person's hot could be 80 degrees when they are used to cold temperatures of minus 30 degrees, whereas another persons "hot" may be 110 degrees, when their "cold" temperature dips below "60." Conceptual mapping may be good for less dynamic situations, however as the speed of change increases, there must be some way to jointly calibrate quickly.

10. (cont)	Perhaps early divergences of team members could provide a means to quickly focus on affordances of importance, and adroitly move the group toward convergence. Individuals allow different affordances to be processed based on current hypotheses and/or focus, using their own personal scales [Kelly's Personal Construct Theory]. Team training could encompass understanding of personal constructs of importance, danger, etc. and could be useful in real-time interactions. Indications of early divergences could be, "To what affordances are team members attending?" Can these actions of attending to these affordances be automatically associated and presented so that patterns of these affordances indicate that the team members are focussing on separate issues? Extending possibilities, can participants be given different coding schemes/indicators based on expertise, closest to the problem, most recent information? For example, the most experienced evaluator, who is closest to the problem, and has the most recent information, his/her actions could be given prominence, indicating affordances that he/she is attending to. Comparisons could be made to provide early detection of divergences. Just as they have found that repeated speech and text patterns by strategic decision makers indicates a shift in attention, similar ways can be found for these shifts in attention. For example, speech and text could be sampled for team interactions in command and control operations and for tactical operations within combat information centers. There may be ways to extend this notion of what affordances in the environment for which one or more team members are searching? and what affordances are strong enough or resonate enough with participants so that they are accepted? For example, technology may be used to map eye glances and head movements. This pattern could indicate a change in attention. In highly complex environments that demand team role differentiation, what is the balance between group cognition and distributed cognition? That is, while there is a need to perform differentiated tasks, there is a need, as has been noted in distributed cognition studies, for maintaining an overlap of role understanding. At the same time, too much attention to the same or similar stimuli, "tunnel vision," may cause inadequate attention to a broad range of stimuli that are to be handled by a team with differentiated roles. As indicated above, this could be applied to command and control and in directing unmanned air vehicles. There are at least two ways where these ideas can affect process and development of support systems. From combat information centers with large groups with differentiated roles to unmanned vehicles where teams are located separately from the vehicle, there are multiple affordances of stimuli that exist in the task environment. Mechanical devices that record head and eye movement, coupled with computers that are analyzing these movement patterns, plus analyzing speech and visually displayed information (including text) could provide early information of divergence of importance and early signs of distributed cognition that might need to be shared with the group. Secondarily, within large and small teams with differentiated roles, there may be coupling of humans and machines that redundantly focus on single roles. In addition to mechanical recording of head/eye movement, speech, and text transmissions, other means that support more intensive creation and maintenance of shared understanding of the situation may prove useful. Any means that provides early warning of divergence and a simple means of calibration of shared meaning could be especially useful.

11. Time anonymity and open conversation: Support security and procedural protocols that provide opportunities for this.	For example, anonymously, one person can request during times of topic exploration that the discussion be anonymous. If this is allowed in the protocol for the task and the group, or for the phase of the task, then all would be anonymous. Other times, there may be a requirement along some range for a number of people to indicate the desire for anonymity, and then the system could respond
12 Provide means to quickly capture subsets of statements and categorize into domains and attributes of domains – provide relevant domains to ease selection	Wynn [1997]; Boland and Tenkasi [1995].

5. CONCLUDING REMARKS

This paper focussed on identifying multi-theoretical foundations for supporting group sensemaking in the social construction of knowledge and artifacts. Based on these foundations, some initial guidance to augment the social construction of knowledge and artifacts were identified. Because of the breadth and complexity of the subject, this report must be considered a work-in-progress, a snapshot of the exploration of such a complex subject.

6. REFERENCES

Ackermann, F. "Use of Mapping Technology for Strategic Planning," Presentation, Computer & Information Sciences Department, August 1997.

Boland, R. J. and Tenkasi, R. V. "Perspective Making and Perspective Taking in Communities of Knowing," Organization Science, Vol. 6, No. 4, July-August, 1995.

Bostrom, R. P. "Successful Application of Communication Techniques to Improve the Systems Development Proces," Information and Management, Vol. 16, 1989.

Churchman, C. W. The Design of Inquiring Systems, Basic Books, New York, 1971.

Daft, R. and Lengel, R. (1986), "Organization Information Requirements, Media Richness and Structural Design," Management Science, Vol. 32, No. 5, May, 554-571.

Davenport, G. and Murtaugh, M. "Automatist Storyteller Systems and the Shifting Sands of Story," IBM Systems Journal, Vol. 36, No. 3, 1997, 446-456.

Dreyfus, H. & Dreyfus, S. (1986) Mind Over Machine, Free Press, New York.

Edamala, M. "Social Construction of Knowledge and Artifacts: Knowledge Creation," Computer & Information Sciences Dept. Working Paper, December 1997.

Finke, R. A. and Bettle, J. Chaotic Cognition: Principles and Applications, Lawrence Erlbaum Associates, Mahwah, New Jersey, 1996.

Gephart, R. P. "The Textual Approach: Risk and Blame in Disaster Sensemaking," Academy of Management Journal, Vol. 36, No. 6, 1993, 1465-1514.

Hedberg, B. and Jonsson, S. "Designing Semi-Confusing Information Systems for Organizations in Changing Environments," Vol. 3, No. 1, 1978, 47-64.

Kelly, G. A. A Theory of Personality (2 vols), Norton, New York, 1955.

Larkin, J. H. and Simon, H. A. "Why a Diagram Is (Sometimes) Worth Ten Thousand Words," in Diagrammatic Reasoning: Cognitive and Computational Perspectives, Glasgow, J.; Narayanan, N. H.; and B. Chandrasekaran, eds., AAAI Press / The MIT Press, Menlo Park, California, 1995, 69-109.

Moray, N. "Intelligent Aids, Mental Models, and the Theory of Machines," in Cogntive Engineering in Complex Dynamic Worlds, Hollnage, E.; Mancini, G.; and Woods, D. D., eds., Academic Press, New York, 1988.

Multiple Intelligences Approach to Assessment, Zephyr Press, Tucson, Arizona, 1997.

Nah, F. H. and Kim, B. "Influence and Consensus in Group Decision Making Environment," Proceedings of the Americas Conference on Information Systems, August, 1997, 903-905.

Nonaka, I. "A Dynamic Theory of Organizational Knowledge Creation," Organization Science, Vol. 5, No. 1, 1994, 14-37.

Nonaka, I. And Takeuchi, H. The Knowledge-Creating Company, Oxford University Press, New York, 1995.

Nosek J. T. "Exploring Task Atomicity and a Theory of Group Flow Using Collaborative Programming," America's Conference of the Association of Information Systems (AIS) Proceedings, August, 1997.

Nosek, J. T. and McNeese, M. "Augmenting Group Sensemaking in Ill-defined, Emerging Situations: Experiences, Lessons Learned, and Issues for Future Development," Information, Technology, and People, (in publication), September, 1997.

Nosek, J. T. and Sherr, D. "Getting the Requirements Right vs. Getting the System Working - Evolutionary Development," in Bemelmans, T.M.A. (ed.) Beyond Productivity: Information Systems Development for Organizational Effectiveness. New York: North Holland, 1984.

Petzinger, Jr., T. "The FrontLines," The Wall Street Journal, January 9, 1998, B1.

Preece, J.; Rogers, Y.; Sharp, H.; Benyon, D.; Holland, S.; and Carey, T. Human-Computer Interaction, Addison-Wesley, Reading, Mass., 1995.

Rasmussen, J.; Pejtersen, A. M.; and Goodstein, L. P. Cognitive Systems Engineering, New York: John Wiley & Sons, 1994.

Sage, A. P. "Behavioral and Organizational Considerations in the Design of Information Systems and Processes for Planning and Decision Support," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-11, No. 9, September 1981, 640-678.

Schmidt, J. and Bannon, L. "Taking CSCW Seriously: Supporting Articulation Work," Computer Supported Cooperative Work, Vol. 1, 1992, 7-40.

Schwenk, C. R. "Strategic Decision Making," Journal of Management, Vol. 21, No. 3, 1995, 471-493.

Senge, P. Sharing Knowledge, Vol. 14, Issue 11, 1997.

Shaw, M.L.G. and Gaines, B. R. "Knowledge Support Systems for Constructively Channeling Conflict in Group Dynamics," in AAAI-94 Workshop on Models of Conflict Management in Cooperative Problem Solving, Seattle, Washington, August 4, 1994

Sproule, J. M. Argument: Language and Its Influence, McGraw-Hill, New York, 1980.

Weick, K.E. The Social Psychology of Organizing, 2nd edition, Addison Wesley, Reading, Massachusetts, 1979.

Winograd, T. and Flores, F. Understanding Computers and Cognition, Addison-Wesley, New York, 1987.

Woods, D. D. "Commentary: Cognitive Engineering in Complex Dynamic Worlds," in Cogntive Engineering in Complex Dynamic Worlds, Hollnage, E.; Mancini, G.; and Woods, D. D., eds., Academic Press, New York, 1988.

Wynn, E. "The Social Construction of Knowledge in Information Systems Development," Computer & Information Sciences Colloquium, Temple University, September 30, 1997.

THE HUMAN-ELECTRONIC CREW: HUMAN-COMPUTER COLLABORATIVE TEAMWORKING

R.M. Taylor
DERA Centre for Human Sciences
F138 Bldg
Farnborough, Hants GU14 0LX
United Kingdom

Dr J Reising
USAF Research Laboratory
Wright-Patterson AFB
Dayton, OH 45433-7901
USA

1.SUMMARY

Advances in artificial intelligence will enable future military aircraft to have a rather unique crew - one human and one electronic. It has proved useful to consider the required relationship as a Human-Electronic Crew team, involving collaborative, co-operative working between the human and the machine. This team is led by the pilot, with the Electronic Crewmember as a subordinate associate or assistant, sharing responsibility, authority and autonomy over many cockpit tasks. As aircraft systems become more complex, the automation that the aircraft pilot has to interact with is becoming increasingly intelligent and capable. The pilot needs to remain in control of the system in uncertain situations with unpredictable demands, and yet make full use of the aiding that is provided, whilst being flexible and adaptive. The requirement for useful, intelligent aiding, in a highly dynamic task environment, has led to impressive technical achievements. These include methods for in-flight situation assessment and replanning, cognitive modelling, human intent inferencing and error recognition, and the use of complex knowledge engineering and reasoning logic processes. Providing an appropriate architecture for complex system functioning, where the pilot can trust the Electronic Crewmember with autonomous aiding, but that keeps the pilot in control, presents a continuing engineering challenge.

2.INTRODUCTION

Ever since the movie *Star Wars* showed Luke Skywalker and R2D2 teaming up to destroy the Death Star, there has been considerable speculation as to how an efficient pilot-robot team could be created. Since weight is a critical factor in airborne systems, the literal building of a pilot-robot team has not been undertaken; rather the emphasis has shifted to incorporating the *intelligence* of the robot. The early vision was of crew-adaptive cockpit, with sensors monitoring the pilot's state, artificial intelligence (AI) software enabling the computer to learn, and cognitively compatible, pictorial displays allowing efficient presentation of cockpit information to the pilot.[1] In the 1980's, developments in real-time data acquisition, fusion, and processing, and in computer modelling and AI inferencing techniques, such as Expert Systems, Knowledge-Based Systems (KBS), and Neural Networks, began to offer the opportunity for realisation of these ideas. As work in this area progressed, such terms as "electronic crewmember" (EC) and "black box back seater" began to enter the vocabulary of both the crew station and computer software communities.

In 1985, the establishment of the US Pilot's Associate Programme gave credence to the idea that the building of the brain of R2D2, in some very simplified form, might be possible. Some of the results of this programme have been transitioned to the US Army's Rotorcraft Pilot's Associate programme. In Europe, AI efforts have centred around a number of pilot aiding projects. These include the French "Co-pilote Electronique", the UK Mission Management Aid (MMA), and the German CASSY and CAMMA Cockpit Assistant Systems. Recent research has sought to develop a theoretically grounded, human-centered approach for guiding a principled development of intelligent pilot aiding concepts for cockpit automation, such as the UK DERA Cognitive Cockpit programme and the USAF Research Laboratory Adaptive Interfaces project. The purpose of this paper is to highlight relevant aspects of this work as an example of human-computer collaborative performance in a complex operational system.

3. HUMAN-ELECTRONIC CREW TEAMWORK

From the outset, consideration of the human factors implications for workload sharing between the pilot and EC, and in particular the effects on pilot situation awareness (SA), led to some difficult questions being asked. These included questions such as: Is the pilot always in charge? Can the pilot and EC really be called a team? Why do you need a pilot at all? This served to stimulate thinking in the area of *human-computer teamwork* in the cockpit. It was recognised that was being considered was potentially a new form of co-operative or collaborative behaviour. For the purposes of definition, co-operative behaviour can be considered as a form of interaction in which *agents* (i.e. humans or machines) share goals, implicitly or explicitly, and act in concert to achieve them. The agents of present concern are the aircraft pilot and the avionics computer coupled as a co-operating team.

Since 1988, a series of Technical Workshops has sought to address a wide range of issues and topics concerning Human-Electronic Crew (HEC) Teamwork.[2,3,4,5] The issues covered have included problems of technical implementation of AI, through pilot-EC dialogue and SA, to the EC's autonomy and building trust between the two crew members, the level of confidence that higher authorities can have in the decisions and the resulting actions of the pilot-EC team. As the concept approaches maturity, sufficient for technical implementation in-flight, the most recent HEC Workshop, held in Kreuth, Germany 23-26 September 1997, sought to focus on the nature of the EC when finished, with the question "The Human Electronic Crew: The Right Stuff?" Specific issues

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

included: What are the key components that will ensure a successful emergence of this technology? How can we plan for their development and incorporate the software and hardware functions in concert with one another? What is sufficient functionality within the EC to satisfy the human operator's requirements?

4. PILOT AIDING TECHNOLOGIES

Collaborative working between the pilot and cockpit aiding can be considered in terms of the capability maturity of the technology. Fundamentally, automation has been designed to replace human control and decision making. But operationally, the requirement is for cockpit technology that supports and maintains the necessary human control and decision making in the operation of the aircraft system. Arguably, cockpit automation technology should be judged in terms of the cognitive quality of the necessary pilot involvement in decision making and system control, rather than by the degree of automation autonomy.

Using a cognitive quality framework, some broad distinctions can be made between the capabilities of conventional automation of task execution, decision aids or tools for decision-making, and "intelligent" adaptive aiding systems.

- **Level 1 - Conventional Automation** At the lowest level of capability maturity, conventional automation was designed originally to improve pilot efficiency and reduce workload. It seeks to replace pilot involvement in the execution of highly repetitive, predictable and tedious tasks, which arguably would be more appropriate for machine performance. This 'out-of-the-loop' performance with the pilot in a supervisory and monitoring role, can lead to dependence, complacency, mode awareness problems, and isolation of the aircrew user. This isolation can be dangerous for a pilot with ultimate responsibility for the operating safety of the aircraft. Another problem is that the use of highly sophisticated control systems has expanded the role of automation to include tasks that blur the distinction between decision making and task execution.

- **Level 2 - Decision Aids** Decision aiding has followed a different developmental path without seeking to replace human involvement. Decision aids seek to provide tools to assist the user in organising data, structuring alternatives and evaluating options. But, by adopting an optimisation strategy, which seeks to provide the best decision outcomes, decision aiding has began to resemble conventional automation through progressive automation of the decision process, rather than aiding the human decision maker.

- **Level 3 - Adaptive Aiding** At the highest level of capability maturity, adaptive aiding seeks to intelligently augment and enhance human judgement and responsibility, supporting the human and mitigating against his/her limitations. This be considered as dynamic aiding which operates flexibly, adapting in response to changing requirements; to achieve this, its functioning can be contextualised with regard to both the individual pilot's needs and the dynamic external situation requirements. Adaptive aiding can support the pilot's decision making intelligently by proposing a candidate solution for consideration by the pilot, or it could critique a proposal generated by the pilot. If necessary, it can propose, select and execute the solution for the pilot. Significantly, it can have the ability to judge the level of aiding that is necessary. Intelligent adaptive aiding is based on consideration of human limitations and capabilities, rather than of system and mission performance i.e. it is human-centred and constraint-based. It seeks to restore the pilot to the role of the decision maker, at the same time as providing safeguards for situations in which time limitations, or the complexity of the problem, restrict pilot problem solving ability.

5. INTELLIGENT AIDING SYSTEMS

Adaptive aiding can be considered to be intelligent to the extent that they produce behaviours that have the characteristics of intelligent human-like behaviours c.f. the Turing test. There is broad agreement that intelligent human-like behaviour should have the following features[6]:

- Active collector of information
- Goal driven
- Reasons at multiple levels
- Context sensitive communicator
- Learns from experience.

An intelligent system comprises primitive processes or behaviours, and the knowledge that they manipulate. All gross, higher level behaviours emerge as a result of the interactions between these primitive processes. The system can reorganise its primitive behaviours to adapt to a very large set of situations. A small set of primitive processes, combined with representations of knowledge, can produce a vast set of intelligent behaviours.

On the basis of this understanding, intelligent aiding systems can be considered to have three universal characteristics[6]:

- Emergent properties - They demonstrate emergent characteristics as a result of their interactions between their components, and the functional capabilities of assessment, planning, co-ordinating and acting.
- Functional integration - The desired behaviours for domain requirements are shared across the functional components including the human user. The same function may be performed by more than one functional component.
- Open ended - They are capable of absorbing new knowledge and creating new behaviours without changes to its processing, thereby avoiding obsolescence.

Unlike the tool-set oriented approach of decision aids, intelligent aiding can be considered as providing a highly integrated comprehensive aiding system, intelligently integrating many decision functions to provide an interlocking system. The interactions between the major functional components (assessment, planning, acting, co-ordinating) produce emergent properties that enable complex, flexible and adaptive responding.

Functional integration, rather than function allocation, is an important characteristic of intelligent aiding systems[7]. As tasks become more about thinking than doing - more cognitive than physical in nature - the validity of applying ideas of functional separation and allocation to cognition has to be questioned. Analysis of cognition into separable functions, which become candidates for automation, may be counter-productive. HEC teamwork may benefit from a functional architecture with integration, rather than segregation and allocation, of high level functions. With functional integration, the behaviours required by the domains are shared across the functional components, including the user. Thus, intelligent aiding forms a joint cognitive system with the user. The same function can be performed by several functional components, rather than just one, providing robustness and more operational flexibility than when functions are allocated to specific system components.

Intelligent aiding systems under development can be distinguished in terms of three main types, in accordance with the tasks and roles that they perform, and the knowledge that they manipulate[6], namely:

- Assistants - perform specific tasks when asked, using basic task and situation knowledge
- Associates - recognise that the user needs assistance, using complex task and situation knowledge, and basic user and co-ordination knowledge
- Coaches - both aids and instructs to assist the user better, using complex task, situation, user and co-ordination knowledge

Most intelligent pilot aiding systems currently under construction operate at the level of pilot assistants, but with the potential for the development of functionality needed for higher levels of aiding.

An aiding system can be considered more as an intelligent cockpit that as a conventional cockpit display and control centre[8]. Through the use of knowledge and reasoning processes, the cockpit takes on an agent-like qualities:

- It intelligently responds to user commands and requests, and delivers pertinent information.
- It provides knowledge-based state assessments.
- It provides execution assistance when authorised.
- It engages in dialogue with the crew, implicitly and explicitly using cognitively-based transactions, at a conceptual level of communication and understanding.
- While assisting in mission execution, the intelligent aiding makes the crew station interface more useable and non-intrusive i.e. it provides both useability and mission aiding.

Intelligent aiding systems provide assistance with the basic functions of assessment, planning, co-ordinating and acting. In the cockpit, this translates into aiding with the following:

- situation assessment through actively gathering information on changes and reporting significant events,
- planning and responding,
- management of task and resources
- information and control presentation.

In a multi-crew cockpit intelligent aiding will need to cope with the trading of tasks and responsibilities between the crewmembers, which will produce more complex intent inferencing, information management, error detection and adaptive aiding[9]. Opportunities will arise to facilitate crew co-ordination, such as through multiple task tracking, dynamic task allocation to balance task loading, and monitoring task performance and sequential interdependencies to warn or aid when dependencies are in danger of violation.

The potential for co-operation through intelligent aiding is achieved by:

- incorporating a model of human decision making and control abilities into the control automation,
- monitoring pilot performance and workload through behavioural and physiological indices,
- predicting pilot expectations and intentions with reference to embedded knowledge of mission plans and goals.

Functional architectures are required for intelligent systems that support strong interactions and tight integration. Candidate architectures include hub, layered and federated structures. A generic, reusable architecture for a platform-oriented system has been proposed[6]. This is based on a multi-tasking, multi-processing executive with assessor, planner, intent model, and execution aid and information manager modules. These are linked to controls and display functions and to mission equipment and platform systems through a data distribution system. The architecture should have scale-up potential for the integration of multiple co-operating intelligent aiding systems, sharing data with a large macro system.

Figure 1 shows a simplified architecture for the essential adaptive cockpit interface components, suggested by the USAF Research Laboratory, and also used to guide the DERA Cognitive Cockpit programme. This couples a situation assessor (probably KBS) and a pilot cognitive state estimator capability (comprising behavioural and physiological measures), through adaptive algorithms in the Pilot Vehicle Interface (PVI), driving adaptable, modifiable PVI devices. The resultant PVI adaptations could take the form of display modifications that deliver more readily interpreted information, warning messages, correction displays and augmented feedback control feedback, all of which leave the pilot fully in command of the aircraft.

6. ASSESSING TEAMWORK

Can formal methods be used to assess successful pilot aiding and HEC teamwork? Developmental frameworks can characterise progress in aiding technologies and provide methods to measure progress. Formal methods can enable prediction of the effectiveness or accomplishment for pilot aiding technologies, and their capability maturity.

Trustworthiness is important characteristic of aiding. Passing the Turing test (i.e. successfully mimicking human-behaviour) may not be sufficient. For trustworthy collaborative working, the aiding needs to be both consistent and correct i.e. non human-like[10].

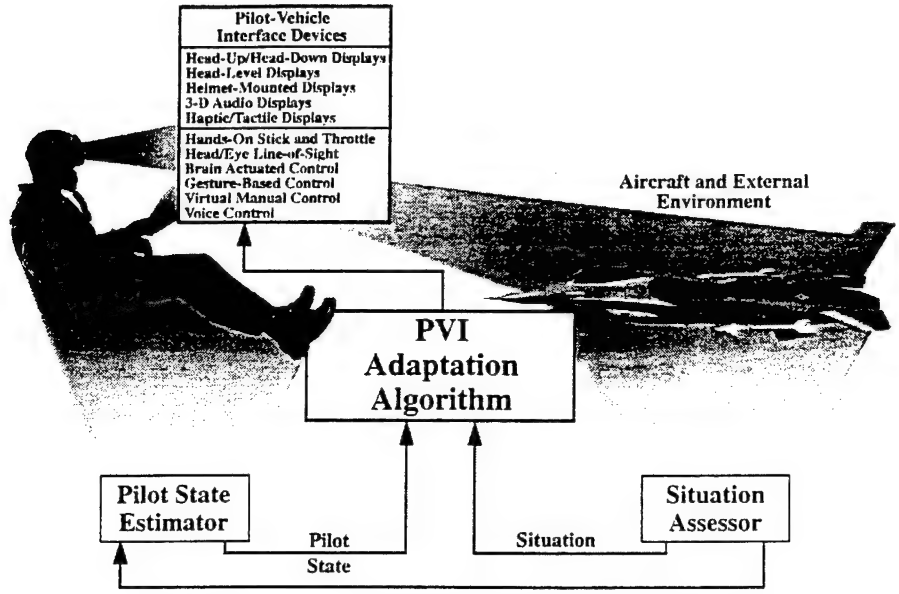


Figure 1. Adaptive Cockpit Interface Components

The required framework probably needs to be expressed in terms of both the decision-making nature of work and of the nature of the supporting technology. A complex Work Process Framework has recently proposed[11], based on cognitive systems engineering (Rasmussen's SRK paradigm). This approach highlights aiding opportunities whilst preserving the fundamental continuity and flow characteristics of work.. It comprises an abstraction hierarchy that captures means-end relationships (function; information; physical processes; form and configuration) and a task work decomposition that distinguishes between procedural activity, embedded problem solving, and off-line problem solving activity. It also includes an SRK model with an iterative understand-act performance loop. The need for different assessment perspectives is recognised, with multiple levels of abstraction. One example is a human-by-interface perspective matrix, with the dimensions sensory/perceptual-cognitive support, understand-execute, and interface content-form-behaviour.

Typical decision aiding taxonomies (e.g. Sheridan and Verplanck[12]) assign behavioural elements or cognitive functions (e.g. perceive, decide, execute) to levels of automation (e.g. decision aid tool, decision support assistant, associate of the human, fully autonomous agent). Table 1 provides a simple example of the balance of responsibilities. More powerful analysis can be made by focusing on the knowledge-based requirements for supporting decision-making, considering levels of decision automation in terms of the completeness of knowledge of the characteristics of decisions (e.g. goals, world knowledge, action options, actions outcomes, desirabilities)[13].

PILOT AIDING TECHNOLOGY	PERCEIVING	DECIDING	ACTING
Conventional Automation	*PILOT*	*PILOT*	*EC*
Decision Aids	*PILOT*	*PILOT-EC*	*PILOT*
Adaptive Aiding	*PILOT-EC*	*PILOT-EC*	*PILOT-EC*

Table 1. Simple Pilot Aiding Taxonomy

HEC TEAMWORK STRUCTURE - CRITERION DIMENSIONS			
GOALS	**RESOURCES**	**STRUCTURE**	**PROCESSES**
Clarity - defined performance objectives **	**Sufficiency** - enough expertise & ability**	**Goal driven** - governed by performance	**Wide bandwidth** -many comm. modes
Common structure - shared understanding**	**Availability** - readiness for application*	**Accessibility** - facilitates access to resources*	**Bidirectionality** - two way information flow**
Tracking - awareness of changing objectives*	**Heterogeneity** - variable/unique expertise	**Cohesiveness** - attracts conformity to norms*	**Shared initiative** - leadership turn-taking*
Impact - critical for mission success*	**Compatibility** - ability to combine & integrate**	**Dynamic Function Allocation** - real-time role & task distribution**	**Common knowledge**- shared understanding**
Achievement - high probability of success**	**Enhancement** - ability to add expertise	**Levels of Autonomy**- degrees of independent functioning	**Trust** - willing to accept others' judgements**

* Moderate sensitivity ** Strong sensitivity

Table 2. HEC Teamwork Dimensions

Another perspective considers the nature of teamworking. The literature on teamworking and on pilot aiding has indicated that teams have three distinctive characteristics.

- Co-ordination of activity, aimed at performing certain tasks and at achieving specific, agreed goals. Such co-ordination is dependent on trust between team members to be successful, since trust is the mechanism which allows co-ordination of effort to take place.
- Well-defined organisation and structure, with members occupying specific roles with associated power, authority and status, whilst exhibiting conformity and commitment to team norms and goals. Such organisation will define the allocation of functions and the locus of authority within the team.
- Communication and interaction between team members. These are referred to as team processes.

A framework model based on this analysis has been used to measure and guide the development of teamwork in pilot aiding[14]. The framework proposes key requirements for team goals, resources, structure and processes, and candidate constituent dimensions, as the basis for auditing teamwork quality (Table 2).

Some evidence is available to judge the validity of the teamwork perspective approach. Validation studies have investigated the sensitivity of the framework dimensions and their diagnostic power. In one study, ratings of the teamwork dimensions were made by experts comparing working with immature and mature aircraft and support systems in different operational roles[14]. The results indicated some sensitivity to expected operational role and technology maturity differences. The importance of a good human-computer interface was highlighted. A follow-up validation study compared the Tornado GR1 and advanced Harrier GR7 cockpit[15]. Teamwork ratings were provided by experienced aircrew for typical stressing scenarios of pilot-technology teamworking (good and bad). The results showed strong sensitivity to the scenario teamworking differences for nine of the dimensions, and moderate sensitivity for six dimensions, as illustrated in Table 2. The aircraft types differed only in terms of the assessed team processes. The quality of communication, initiative and trust was rated higher in the two-crew Tornado cockpit, compared with the single pilot Harrier cockpit. If two-crewmember communication processes are a quality benchmark, or goal, single-seat HEC team processes are relatively poor. Bilaterally initiated communication, intent inferencing and trust are important characteristics the of Tornado cockpit communication[17].

A third validation study of the teamwork framework reported ratings of simulated flight task aiding using adaptive automation[18]. The results showed relatively high ratings on the goals dimensions, but low ratings of the resources, structure and processes. Thus, while the aiding concept aim was assessed as good, its implementation in the simulation software was considered poor (i.e. invocation logic, levels of automation, status indications, interface interactions). Mode awareness problems with flight deck automation ("What's the kit doing now?") indicate that poor human interfacing with cockpit automation is a general problem.

Team function allocation may be appropriate for controlling systems with discrete, bounded, naturally separable functions and tasks, where any resultant autonomy only weakly threatens goal maintenance. With ill-structured problems involving uncertainty, function allocation needs to be flexible and dynamic, to reduce dependency, with good communication between team-members. Mature human teamwork involves good communication, function and leadership initiative turn-taking, with a transitioning of authority that is smooth and

flexible. The locus of control is driven more by situation and context, than by the preservation of a sole source of control authority.

7. OPERATIONAL PHILOSOPHY

Many of the key questions and issues concerning HEC teamwork lead to consideration of the required operational relationship between the pilot and the EC. Logically, the operational philosophy, or concept of operation of the system, should provide the rationale for determining the roles of the pilot and the EC in the system functioning. In a human-centred approach, the desired role of the pilot should determine the nature of the required operational relationship between the pilot and the EC. From understanding of the pilot's role within the system, the nature of the system support required for that role can be determined, and this analysis can then guide the system design.

Must the pilot be a human role? Why do you need a pilot at all? In the most stressing case, the nature of military conflict dictates that the pilot is expected to operate more effectively than adversaries in uncertain and extreme circumstances, at the limits of knowledge, skill and performance, whilst considering complex interactions, with significant consequences, and with considerable freedom of action. Involvement of human intelligence in decision making seems to be particularly important, and probably essential, in solving complex real-world problems. The problems of greatest concern - adversarial conflict, aircraft emergencies - are mostly situations which are unstructured, involving uncertainty and incomplete information, with multiple, competing, and unstable goals. Human intelligence is needed because it is not constrained by formal logic rules or probability. Humans can make leaps of abstraction and intuition, producing new solutions to novel problems. This often makes human intelligence superior to machine intelligence, when the need to adapt and learn is critical[19].

From understanding of human role requirements at the highest level, such as the pilot's needs for system control, analysis can indicate the human limitations that apply to the role, such as cognitive issues of motivation, attention, skill and knowledge. Requirements analysis can then identify alternatives for dealing with the expected human limitations, and for enabling the required performance and control to be achieved, with the assistance of aiding technology.

Given the pilot's difficult role, and consideration of the required HEC operational relationship, the building of trust between the pilot and the aiding system, and between the HEC team and outsiders, is a key issue. Trust is built when consistency and correctness is observed in the team's decisions and actions. Two important guidelines for building trust have arisen[10].

- Define EC's Prime Directives. These are overall governing rules which bound the behaviour of the EC, and yet provide a logical structure for EC to act in a rational and reliable manner, avoiding arbitrary behaviour, so that the pilot does not experience any surprises e.g. Asimov's Laws of Robotics.
- Specify EC's Levels of Autonomy. These also bound the behaviour of the EC by limiting its decision authority for the performance of specific sub-functions to a set of system configurations specified and set by the pilot.

Comparisons between the approaches taken by the Co-Pilote Electonique, CASSY/CAMMA and Pilot's Associate projects reveal some important differences in philosophy and implementation strategy.

8. ASSISTANT SYSTEMS

In Europe, the French Co-pilote Electronique (CE) project and the German Cockpit Assistant System (CASSY) programme, aim to provide cockpit assistant systems, with limited automation autonomy, intentionally short of a full associate relationship.

CASSY is a civil aviation cockpit assistant project. As such, it has been guided by the principle concern of flight safety[20]. The top level, human-centred requirement, it is that successful automation should be designed to avoid situations in which the crew is overloaded. Thus, the aim is to keep crew workload demands at a normal level in all situations and in their constituent tasks. Two basic requirements have arisen from this:

- In presenting the flight situation, the system must ensure that the attention of the crew is guided to the most objectively urgent task or sub-task.
- If the above requirement is met, and the crew are overloaded, then the situation has to be technically transformed into a situation that can be handled normally by the crew.

This approach has led to a strong emphasis in the CASSY project on assisting the pilot with situation assessment and replanning in flight. Situation-dependent assistance with flight planning is guided by a normative pilot model, goal conflict and intent/error recognition, with aiding in the execution of pilot selected functions. When the need for plan modification is recognised from conflicts detected in the situation evaluation, a new plan is generated and recommended to the pilot. If the pilot rejects the plan, the next best alternative is offered. The pilots actions are monitored for the expected actions needed to follow the new plan. If a discrepancy occurs between the expected actions and the actual aircraft state, a synthetic speech voiced hint is given to the pilot to enable early correction of small errors, and avoidance of developing safety hazards[21].

The French CE project aims provide cockpit assistance to the military fast-jet combat pilot. It is guided by a human-centred, top-level operational philosophy[22].

- The pilot needs to anticipate situations and problems - they need assistance with anticipation and planning ahead, using a 'what if' approach and less reactive aiding.

- Pilot's decisions often compromise between the mental load and the ideal response - the assistance need not be optimised if there is insufficient time for understanding.
- Pilots organise work differently in accordance with individual skill levels - the assistance must be adapted to those skills.
- Pilots want consistency, and do not want to be surprised - the assistance should use a constant model based on pilot logic for understanding and interpreting situations.
- Pilots want assistance that knows and respects its own limits.
- Pilots want a dialogue with assistance that is adapted to the context, to the pilot's intentions, and to the pilot's load.

This philosophy has led to a set of ergonomic rules governing the implementation of a CE intelligent decision support system:

- Homogenous style
- Experience level
- Metaknowledge of competence
- Support anticipation rather than reaction
- Advice dependent on time pressure
- Respect pilot response to incidents
- Ecological interfaces
- Exact information for specific actions, and understanding according to mental models
- Adapt information to intentions

This analysis has led the CE project to take the approach of providing cognitive assistance with situation assessment and planning, but not plan execution[23]. Intent recognition and intent planning are performed to avoid adding to pilot workload. Assistance is provided in situation assessment and planning reasoning by multi-agent expert modules in the domains system management, tactics management and mission management. These modules communicate using a plans-goals exchange language, and work through a co-ordination supervisor. Reasoning about planning directives, restricted action choices to and their consequences considers problem scope, constraints and pilot strategies, adapted to the current situation and pilot mental load. The dialogue with the pilot is handled by the supervisory expert function. This ensures that a single coherent proposal is offered by the domain experts, to minimise information complexity and pilot workload.

The CE project is targeted on the Rafale, with a 2010 horizon. It is currently in the software development phase. Non real-time, full mission simulation demonstration is anticipated soon of dynamic replanning proposals, leading then to refinement of real time constraints. The CASSY project is particularly significant since it has progressed successfully through real-time demonstration and flight test, and it has led to the CAMMA military cockpit assistant project.

9. ASSOCIATE SYSTEMS

The DARPA/USAF Pilot's Associate (PA) programme, 1985-1992, is the foundation project in this area. The PA programme and philosophy continues to exercise influence through current US Army and USAF pilot aiding projects. It should be regarded as indicative of an evolving philosophy, rather than as a definitive understanding of the required system design.

Awareness of the needs of human requirements analysis, led the PA design to be guided by a top-level operational philosophy[24,25].

- The pilot is in charge - i.e. the pilot shall always have the capability to act according to his desires.
- PA's plans may be:
- Approved or rejected explicitly with little effort
- Approved or rejected pre-mission
- Approved or rejected implicitly by pilot action, or
- Ignored with predictable results
- The effort required of the pilot to control the PA must be less than the effort saved by the PA. PA shall save more effort for the pilot than it creates - it shall be responsive to the pilot and not demanding of his resources.
- The PA must operate in a predictable manner.
- The PA is required to monitor the pilot, not the other way around.
- The PA must notify the pilot of key mission events (as defined and set by the pilot).

The goal of the PA was to provide consistently correct information, and to aid the pilot's decision making by helping to manage workload, reduce confusion, and simplify tasks. This led to the philosophy of the EC as an intelligent subordinate to the pilot, with specific capabilities, namely[26]:

- PA could not act on its own.
- PA could make recommendations.
- PA could take actions based on pilot discretion.
- PA could take action based on interpreting pilot intent.
- PA could fly the aircraft tactically on autopilot.
- PA could deal with ambiguities in human speech in the context of the mission.
- PA could diagnose malfunctions, identify miscommunications, and determine the correct response.

These top level requirements led to specific operational relationships (ORs) for discrete EC subfunctions interactions, with increasing degrees of automation and autonomy[27]:

- OR1. The pilot must perform the activity
- OR2. The activity is performed automatically by the EC
- OR3. EC may remind the pilot, if the pilot asks, or has authorised such.
- OR4. EC may remind the pilot.
- OR5. EC may prompt the pilot.
- OR6. EC has been given authority to perform, but with pilot consent.
- OR7. EC has may perform an action only if various conditions are met.

From these ORs, pilot selectable levels of autonomy (LOA) were obtained for groups of functions governed by the required pilot operational relationship and interaction. Discrete LOA modes have been proposed (Inactive, Standby, Advisor, Assistant, Associate), with tailorable functional clusterings for flexible responding, to avoid too rigid automation imposed by

design[28]. In Assistant mode, the PA would maintain advisory functions and also assume responsibility for tasks explicitly allocated to it by the pilot. In Associate mode, under full dynamic function allocation (DFA), the proposed system maintains advisory functions and accepts pilot allocated tasks, but also takes over tasks as the context demands. These modes aim to provide bounded, communicable structure for delegated levels of authority, minimising mode confusion, and building trust and confidence.

Human factors research indicates that the required control structure should be cognitively simple, and not complex. Pilots tend to view EC autonomy simply as either automatic, with or without status feedback; semi-automatic, telling what will happen and asking permission to proceed; or advisory, providing information only[29]. Aircrew recognise DFA (real time role/task distribution), but not LOA (degrees of independent functioning), as important for teamwork quality with current systems[15].

The PA "associate relationship" has been be characterised as a mixed-initiative approach to collaborative problem solving between one or more human actors and a subordinate semi-autonomous computer system with sufficient depth and range of intelligence to encompass a full task domain. The domain knowledge acquisition task, and the knowledge engineering problems of co-ordinating with multiple domains, present significant challenges for building associate systems[9].

An associate system can be characterised in terms of its completeness, and its relationship with the human operator[6]. Through its completeness, the following apply to the associate system:

- It provides complete task coverage within a domain,
- It addresses all areas of human physical and cognitive limitations that are encountered
- It allows the associate to perform any task that the human can perform
- It is capable of fully autonomous behaviour

The associate's relationships with its human operator are as follows:

- Its functions are not allocated exclusively to the human operator or to the associate
- Its functions are allocated both to the human operator and to the associate
- The associate performs tasks only when so authorised by its user
- The user is always in charge of the associate
- The associate must adapt it's behaviour to meet the desires of the user, not the other way round
- The associate's capability extends only to those tasks that have been permitted by the user

Whilst the DARPA/USAF PA programme ended in 1992, it has provided foundations for the US Army's RPA project, providing a two-crew member Cognitive Decision Aiding System (CDAS), which is currently approaching simulation and flight test. The USAF SBIR Hazard Monitor knowledge-based system designed for system malfunction management in transport aircraft is a further development[30,31].

Differences in the planned implementation of pilot aiding between the CE and PA programmes can be identified arising from operational philosophy and technical strategy. The CE project emphasis on pilot involvement and judgement may also reflect a different approach to pilot training, judgement and skills. Peter Svenmarck, at the Swedish Center for Human Factors, Linkoping, Sweden (1998, personal communication) has produced an analysis that highlights the CE emphasis on supporting the individual pilot's problem recognition and situation assessment, and the PA focus on supporting problem analysis and the generation of solutions (Table 3).

The PA objective of a full associate relationship has yet to be fully realised. The USAF SBIR Hazard Monitor implementation currently focuses on the bounded domain of systems malfunctions. The RPA project seeks to provide a Cognitive Decision-Aiding System (CDAS) and a Cockpit Information Manager (CIM), but the functionality and capability is not finalised. Compared with the original PA programme, the CE project, and clearly the CAMMA project, are less ambitious, but probably more technically realistic. But, both are development programmes, aimed at incrementing the technical implementation of concepts and ideas.

10. KNOWLEDGE ENGINEERING

The use by the aiding system of a knowledge-base and reasoning logic processes capable of handling uncertain, imperfect and temporal domain information significantly elevates the level of collaborative co-operation with the pilot. Knowledge engineering is probably the most significant additional cognitive design requirement for cockpits arising from intelligent aiding. An advanced intelligent cockpit could use a knowledge base and reasoning logic processes to provide the following[8]:

DISTINGUISHING DIMENSION	PILOT'S ASSOCIATE	CO-PILOT ELECTRONIQUE
View of pilot expertise	Task analysis	Activity analysis
Cognitive modelling	Plan-goal graph	Can only be understood within combat aircraft
Phase of problem solving	Generate alternatives	Conceptualisation
Philosophy for support	Allocation	Self-reflection
View of pilot error	Correct	Regulation of action

Table 3. Comparison of PA and CE Philosophies

- Intelligent responses to use commands and requests
- Knowledge-based representation and state assessments
- Reasoning for abstract context-sensitive understanding
- Reasoning about system and user constraints (capabilities and limitations) on performance
- Execution assistance when authorised
- More usable and non-intrusive interfaces
- Transactional cognitive-based dialogue interactions with the pilot

Traditional approaches to requirements capture, such as pilot questionnaires and interviews, are not sufficient to provide implementable and safe pilot aiding[32]. Experience has shown that the application of knowledge-based systems (KBS) to real-time decision support systems requires the use of specialised Knowledge Acquisition (KA) methodology and tools, designed to understand how expert knowledge is structured and utilised e.g. DEMUSE, KADS, REKAP, CDM and PC-PACK[33,34].

Cognitive design requirements for an advanced knowledge-based aiding system need to include all the system factors that are essential for the system to behave at a conceptual (abstract, symbolic) level, and engage in a knowledge-based discourse with the system user. This includes knowledge domain requirements arising from the mission and the external environment, and from the system and user capabilities and limitations.

Different of knowledge-base and reasoning structures can be identified for useability and mission aiding[8]. But, understanding of the capabilities and limitations of the human user are key cognitive requirements for both useability and mission aiding. The major challenge for intelligent aiding is how to account for and predict the form of adaptive behaviour that will be exhibited in a given task context.

The required orientation of KBS support for pilot intentionality might be characterised better as constraint-based, rather than as goal-based pilot aiding. For real-time pilot decision support, in complex, dynamic situations, the important uncertainties, and hence the most re-useable knowledge, concern the scope and limits on pilot options, alternatives and flexibility for achieving goals, rather than the structure of the goals and plans per se.

11. INTENT AND ERROR RECOGNITION

A common general guiding principle, on which both the pilot's associate and cockpit assistant philosophies agree, is that intelligent pilot aiding should be designed to support pilot desires or intentionality i.e. goal-based support. For purposes of definition, intentionality can be regarded as a description of an internal mental, or cognitive state involving a focusing of effort and attention on the real world, with a high level of situation awareness, and with a desire, plan, or purpose of achieving some externally reference object, or goal.

The implications for the required pilot-EC operational relationship in an intelligent cockpit are that the EC should not block, or inhibit pilot intentionality. On the contrary, EC should seek to remove barriers and liberate pilot intentionality. This has implications for the form and content of pilot aiding, its invocation and intervention strategy, and significantly, for when assistance is not needed.

To support intentionality, the aiding system must have some means of recognising the pilot's changing intentions and goals. Intent inferencing is a necessary to provide contextualised aiding sensitive to changing plans and goals. Intelligent aiding needs to be able to distinguish between unexpected pilot actions, or failure to produce anticipated actions, which are either:

- errors, arising from human cognitive limitations (slips, lapses and mistakes), requiring pilot advising, cautioning, warning and recovery actions, or
- intentional, and correct, arising from changes in the plan, requiring modification to the aiding provision.

Pilot error can be viewed as action requiring correction, or as normal regulation of action. Because the military aviation environment is often complex and dynamic and unpredictable, pilots require considerable flexibility of action. With uncertain and imperfect data, the true situation may not be well represented within the computer recognised assessment. This presents problems for computer error recognition and classification. Pilot errors are often tolerated deviations of normal control processes for regulation of action. Only errors which are predicted to increase risk of severe consequences without corrective action, are likely to require assertive intervention and action aiding; advisory messages and hinting are sufficient for most recognised pilot error situations.

The plan-goal graph (PGG) modelling approach has been developed to address the problem of intent inferencing[25]. It is based on understanding of the pilots actions and inferred intent, provides a framework to co-ordinate PA actions and resource usages, and to explain the PA's recommendations. The PGG approach is a hierarchical task network planning method. It provides a directed cyclic graph of nodes and links that represents a hierarchy of goals, plans and actions, with a specific set of structural relationships. (The PGG nodes are generalised knowledge comprising attributes, both updateable and instance defining, whose values are not yet bounded. Plan nodes represent an abstract set of activity or operations that may or may not have a specific sequence of action defined for it. The parent goal nodes represent the intended effects of the plan node operations. The actions can be performed by an agent, either the pilot or the PA). The PGG modelling method is important because it provides the following[6]:

- intent recognition by differentiating the goals from behaviour,
- representation at multiple levels of abstraction, enabling reasoning to proceed without unnecessary detail,
- representation of alternative responses to circumstance, anticipating flexibility.

A Pilot Intent and Error Recognition module is used in the CASSY cockpit assistant system[35]. Actual pilot activities are compared with expected activities according to the actual flight plan. When deviations occur, a classification process seeks to recognise possible crew intentions, using an inference algorithm based on known intent hypotheses. A priori

probabilities are modified according to actual pilot actions, and the most probable hypothesis is selected to infer either plausible pilot intent requiring dialogue, or probable pilot error, requiring hinting or warning.

In planning the RPA programme, it was recognised the helicopter domain presents different challenges arising from the less sequential and less scripted nature of the tasks[9]. This was considered to have implications for intent inferencing, making the provision of context sensitive aiding more difficult, than for a fast jet, where work is more scripted. Currently active plans are more difficult to track. This favours a goal-based aiding approach. Also, the decreased sequentiality makes it possible for the pilots to communicate their intentions directly to the associate, using a goal-based vocabulary of "intents". The crew would be able to declare and set goals for the associate, and authorise the associate to determine and enact courses of action. A further problem is that human limitations may mean that, at any given time, the number of plans and goals that can be active may be less than those that can be enabled i.e. conditions are satisfied for them to be enacted. So, support for enabled plans may differ from active plans in terms of different information requirements. An enabled but not active plan may need information about success or failure conditions of the plan and its temporal criticality, whereas an active plan needs information for plan execution.

12. COGNITIVE CONTROL

Pilot's want to remain in charge. They want to remain at the top of the system control hierarchy. They want to be in control of the system, rather than to be controlled by the system. Diminished control means increased unpredictability for the human operator. The problem is how to enable the pilot to have the required control of the system, without being overwhelmed with system control information.

One recent approach sees the pilot's need to be in charge, as a basic sociological problem, affecting the implementation of associate systems. The concept of "tasking" interfaces is introduced to address the problem[36]. This recognises that humans often interact with intelligent subordinates by tasking them to do jobs. Thus, "tasked" systems are always subordinate, but they know enough about the domain that instructing them is vastly easier than instructing traditional automation systems. The automation is given the same task and goal understanding as the human. When combined with a planner, the resulting system permits "tasking" at all the various levels an intelligent subordinate should be able to accept. Tasking interfaces are a method of allowing the pilot to remain fully in control, yet of enabling almost the full autonomy of the associate to plan and execute a high level task whenever the pilot deems that the level of assistance is appropriate.

The pilot needs to retain the ultimate responsibility for generating, setting and changing the system goals and directives. But to maintain control of dynamic system functioning, the pilot needs awareness of the both the actual and desired system status, understanding of the implications of decisions and actions, and control of goal closure through control feedback information.

Extant theory on the cognitive control of complex systems provides a basis for the principled development of intelligent EC pilot aiding, or "Cognitive Cockpit"[37,38]. Rassmussen[39] has provided an error-based classification of behaviour with skill, rule, and knowledge (SRK) levels of performance corresponding to decreasing levels familiarity with the task or environment, or expertise (Figure 2). This approach recognises that behaviour is driven by both goals and experience. It recognises that humans should be allowed to be flexible and variable, and that error observability and reversibility are important features for safe task and system design. It is not clear how control is passed between SRK levels. To account for this, and the orderliness of human action, beyond a stored programme, Hollnagel[40] considers that control of actions to achieve goals exists on a continuum of modes, with different planning horizons, determined by competence (c.f. experience) and context, rather than procedure. These range from scrambled, opportunistic, to tactical and strategic cognitive control modes with increasing levels of depth in the evaluation of outcomes with reference to goals.

In dynamic systems, the observability of the system state and the possibilities for action affecting the state of the system, are key properties of the system to be controlled[41]. A system may be controlled by feedforward or feedback strategy or some combination. In feedback control, only current information is used by the controller about the actual state of the system. In feedforward control, the controller uses a model of the system to predict its state and to select the appropriate control inputs. Feedback control is effective when there are no feedback delays in the system, when the system changes over time, and no stable model of the system can be constructed. Feedforward control requires that the system is stable, so that the model remains valid. Applications of control theory in the automatic control of systems often rely on the model to produce the actual control inputs, and use feedback information to update the model. Feedback control ids cognitively simpler, and is the preferred mode of control in dynamic decision tasks. This works if the rate of change in the controlled system is slow enough for the feedback information to be processed. Hollnagel[42] argues that if there is too much information for the controller to process, then the response will be delayed and the performance will deteriorate. Humans can use heuristics to gain time but the performance will become less precise. Here, it becomes more important to rely more on feedforward and to anticipate responses. It is important that the joint cognitive system retains control of the process, rather than being controlled by it, and that the required stable equilibrium is obtained by a judicious blend of feedforward and feedback control.

Perceptual Control Theory (PCT) [43,44] provides a further useful theoretical perspective for intelligent EC pilot aiding. The main tenet of PCT is that SA, or the individual's perception of it, is controlled by behaviour. It interprets behaviour as an attempt to minimise the difference between desired and actual SA. Figure 3 shows an SA model adapted from PCT[38]. This Integrated Model of Perceived Awareness

Control (IMPACT) has four main constituents: a Perceived Level of SA [P]; a Desired Level of SA [D]; the User's Behaviour [B]; and the Environment [E]. The model works as a basic closed-loop feedback system with a comparator incorporating the feedback in attempting to achieve the goal state.

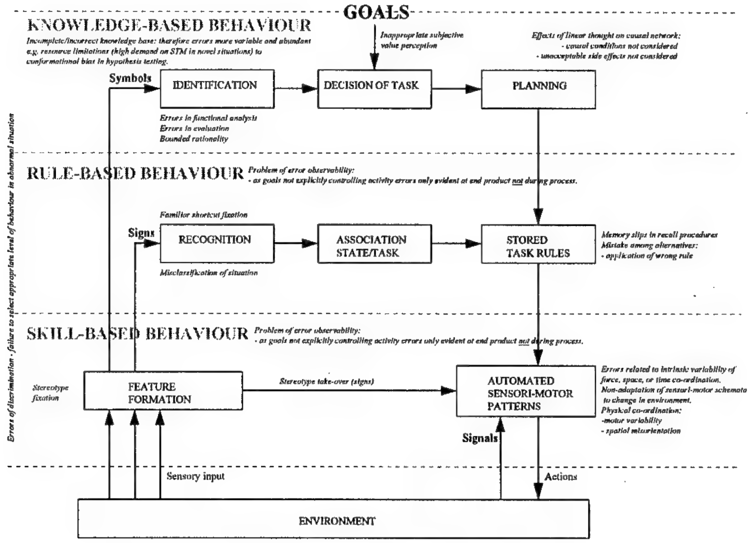


Figure 2. Rasmussen's Skill, Rule, and Knowledge Framework

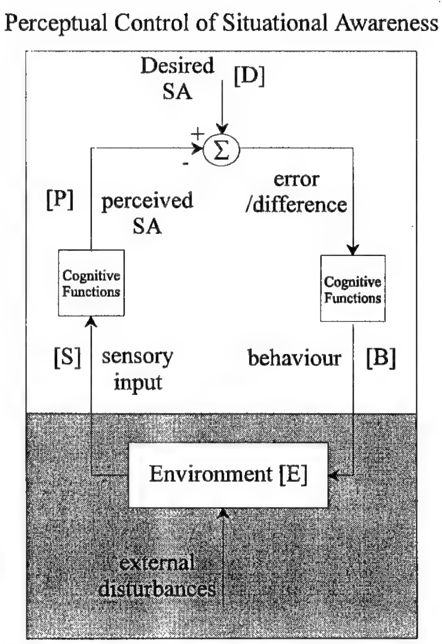


Figure 3. IMPACT Model of SA

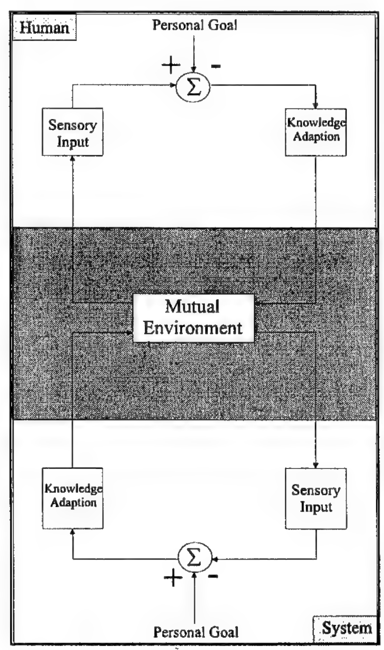


Figure 4. Joint IMPACT Model

There are two main ways the intelligent pilot aiding can use the information the model contains. Firstly, the model provides a reasonable amount of information on the pilot's sensory input. By analysing what perceptions the pilot is taking into account in the decision process, and the quality of these decisions against closure on the goal, the aiding can determine the pilot's informational requirements. Thus, the model can guide communication, and meet the high level requirement of ensuring that the aiding provides an appropriate level of information, of an appropriate quality. Secondly, the model provides some explanation of the pilot's behaviour, in terms of goals, and perceptions. The ability to know why a co-operating agent is performing a certain action, or series of actions, is an essential component of teamwork.

Fig 4 shows how the information exchange in the pilot- EC team can be modelled using the PCT approach, where the EC receives information about the pilot, while it also provides the pilot with information about the EC, and it's actions. A similar symbiotic structure could apply to human-human teams. The model could also incorporate levels of abstraction, with hierarchical control loops corresponding to the SRK taxonomy, modes of control, layers of dialogue, or levels of assertiveness and intentionality. Overall, the main benefit of this model is to guide communication, and provide information about both pilot and the EC. This leads to an increased awareness, which would benefit many complex decision making processes, especially rapid, reactive re-planning.

The DERA Cognitive Cockpit programme seeks to provide an intelligent cockpit system that is sensitive, compatible and adaptive to cognition and to the assessed mission situation and plans, which is supportive to the control of pilot goals, and above all, supportive to pilot intentionality, in accordance with cognitive engineering principles. The intention is that, following the research framework outlined earlier in Figure 1, outputs from a pilot state *Cognition Monitor* (behavioural and physiological) and KBS *Situation Assessor* (and planner) are combined by an *Adaptive Integrator* (agent-based solution), in accordance with a normative and individual cognitive models, coupled with error recognition and intent inferencing. The output is used to drive the cockpit adaptable interface technologies, so as to make the cockpit easier to use, and to provide mission aiding. The Cognitive Cockpit uses the SRK cognitive control framework to identify analyse cognitive requirements for pilot aiding at each of level of control abstraction e.g. tutoring systems for skill-based aiding; expert systems for rule-based aiding; knowledge-based systems for knowledge-based aiding[38]. By structuring all automated support using the SRK framework, this ensures consistency and cognitive compatibility in both the invocation and representation of the automation. The co-operative perceptual control model provides principles for supporting goal awareness (current & desired) and error awareness (diagnosis & rectification), tailored to SRK requirements, with consideration of Hollnagel's modes of control. The current conceptual version of the Cognitive Cockpit, illustrated in Figure 5, provides feedback on EC functioning to the pilot through explicit representations of the activity of SRK agents (dynamic adaptive iconics). Feedforward information is provided through schema-based Goal Balls, indicating action-to-goal effectiveness and risk based on the assessed situation and currently active plans (Figure 6). This provides a cognitively compatible, and ecologically valid representation of uncertainty. SystemCrew Balls represent the *supplied* pilot v EC workload against *required* workload. Support assertiveness is tailored for goal closure in uncertainty, using tutoring, expert advisor, and critiquing techniques, informed by the Cognition Monitor pilot state assessments (Figure 7). The intention is intervene with appropriate assertiveness (hinting, influencing, directing, or if necessary, acting), sufficient to overcome cognitive rigidity without substitution by EC mind set. This approach could resolve the conflicting control requirements for teamwork and autonomy with DFA, by developing a view of EC as an integrated extension of pilot cognitive functioning dealing with uncertainty, rather than as an independent cognitive agent.

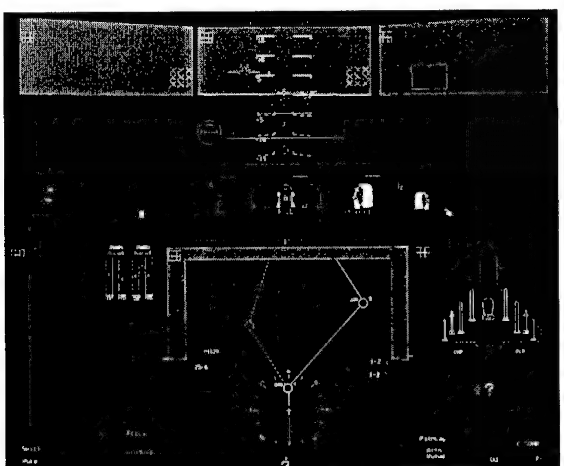


Figure 5. The Cognitive Cockpit

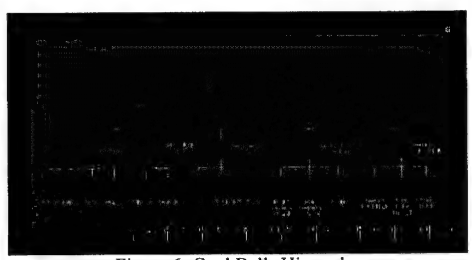


Figure 6. Goal Balls Hierarchy

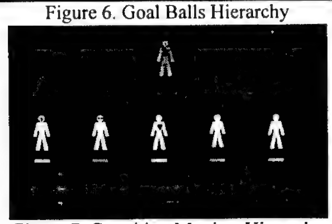


Figure 7. Cognition Monitor Hierarchy

13. PVI ADAPTATION

PVI adaptations associated with assessments of the mission situation and the pilot cognitive state should make the cockpit easier to use, as well as assisting the mission. These adaptations could take the form of display modifications that deliver more readily interpreted information, warning messages, correction displays and augmented feedback control feedback, all of which leave the pilot fully in command of the aircraft. Candidate interface devices for adaptation include the following:

Displays:

- Head-up and Head-down displays
- Head-level displays
- Helmet-mounted displays
- Audio displays
- Haptic/Tactile displays

Controls:

- Hands-On Stick and Throttle
- Head/Eye Line-of-Sight
- Brain Actuated control
- Gesture-based control
- Virtual manual control
- Voice control

Two major questions need to be addressed:

- Under what conditions should interface adaptations be made?
- What specific modifications will enhance pilot performance and increase weapon system effectiveness?

In considering the nature of adaptations that will make the cockpit easier to use, consideration needs to be given to the needs for cognitive consistency and compatibility. Adaptations should not go against the pilot's experience and training. They should conform to the pilot's expectations and not providing unwanted surprises, attracting unnecessarily excessive attention and extra workload. They should be compatible with the pilot's cognitive state, based on behavioural and physiological monitoring, coupled with understanding of human capabilities and limitations in embedded normative cognitive models, as well as geared to the situation demands. Understanding of individual differences in behavioural strategies could also provide useful guidance on the most appropriate adaptation and intervention. Recent research has indicated that some individuals may need support for insight, whereas others need support for responsivity, in situations involving information overload or breakdown of control[45].

A number of relatively robust, and generalisable design guidelines for cognitively compatible interface adaptations can be proposed on the basis of existing evidence on the cognitive quality of pilot interfaces.

Guideline 1: Conform to the Pilot's Mental Model

Mental models play an important part in the efficient operation of systems. Since direct views of the inner workings of a system are often not possible (e.g., the flow of electrons inside the avionics system), displays are a major means of conveying information on the operation of a system. The closer the display formats conform to the pilots' mental model, the more beneficial they will be. Pilots form a mental picture of how a system should work (at a top level) and base their trust in the system according to how the system conforms to this picture or mental model. "A mental model is a <u>representation</u> formed by a user of a system and/or task, based on previous experience as well as current observation, which provides, (most if not all) of their subsequent <u>system understanding</u> and consequently dictates the level of <u>task performance</u>" [46]

Three ideas have been underlined in the above definition to stress its key aspects: representation, understanding and task performance. The pilots' representation leads to their understanding of the system which in turn leads to their performance with the system. For example, if the pilots' mental model of a fuel system pictures the flow valve lever in line with the flow when the fuel is moving and at right angles when the flow is shut off, then that is the way it should be portrayed. It is not important that the valves are electronic and do not have a flow valve handle to turn. An example of not conforming to an operator's mental model is illustrated by the use of reverse notation on early calculators. To add $3 + 2 = 5$ instead of punching the keys in this order, the task had to be performed in the following order: 3 then 2 then +. Needless to say many operators had difficulty in using these calculators.

A cockpit display format called the Pathway head up display (HUD) is an example of matching the display to the mental model because pilots are, in reality, tasked with flying a commanded path in space[47]. Current HUD symbology (Figure 8) requires the pilot to keep the pitch and bank steering bars centred to achieve flight on the commanded path. Conversely, for the pilots to stay on the commanded path when using the Pathway HUD format, they simply fly down the roadway -- a clear mapping of the task and the display used to achieve the task. The roadway is made up of a continuous string of path blocks drawn in perspective, representing 45 seconds of flight into the future (Figure 9). The format incorporates a velocity index displayed in the shape of a small aircraft, called the follow-me aircraft. The follow-me aircraft is drawn to fly along the left side of the pathway at an altitude equal to 150 feet above the desired altitude. It always flies the perfect commanded path at the correct airspeed. To fly the commanded path displayed by the symbology, the pilots only need to fly in an echelon formation on the right wing of the follow-me aircraft. This places the pilots approximately on the centreline of the course. "Road signs" alert the pilots of profile information such as navigation points, glide slope steepness, exact route changes, and a brief description of that change.

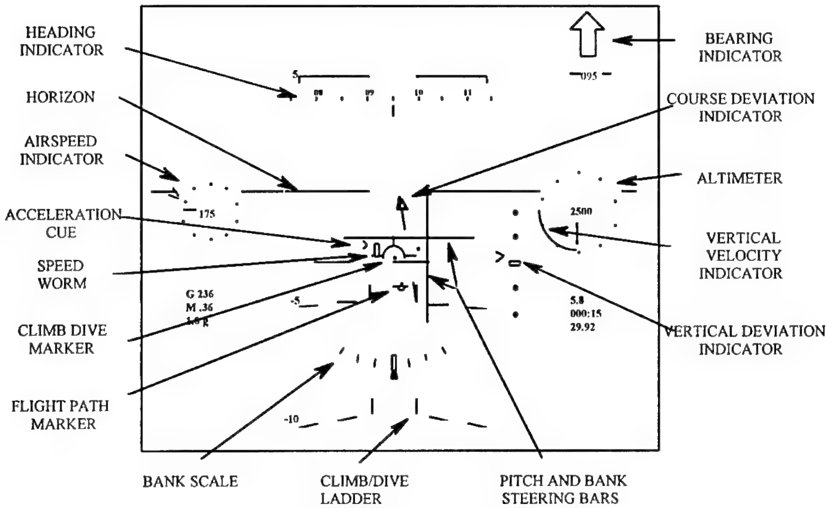


Figure 8 Standard HUD Symbology

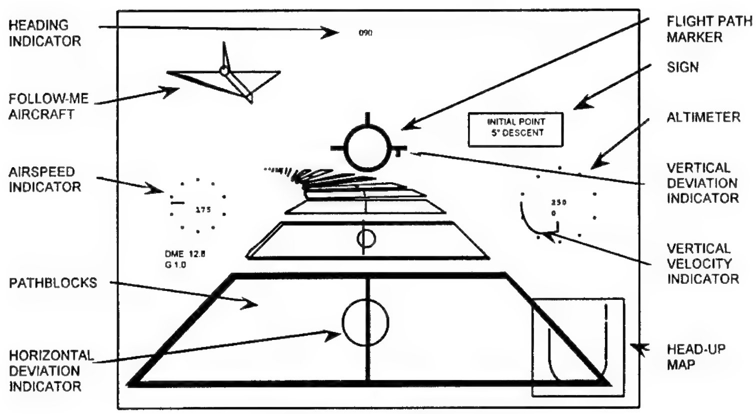


Figure 9 Pathway HUD Format

A comparison between the traditional HUD format and the Pathway HUD format showed that there was a significant difference in pilot performance -- subjects performed better using the Pathway HUD format than the standard HUD symbology in all cases. Pilots comments attributed the advantage of flying the Pathway HUD format to the fact that they could see their route in the form of a roadway from their present position to a point 45 seconds into the future. This allowed them to anticipate necessary control movements. The Pathway HUD format was described as instantaneous situation awareness.

Guideline 2: Make the Interface with the Crewstation Transparent

When pilots communicate with their team members in aircraft such as the Royal Air Force's GR-1 Tornado or the US Air Force's F-15E Strike Eagle, they frequently use voice, a very natural means of communications. Unfortunately, when pilots communicate with the aircraft's onboard computers, they are often forced to wade through numerous levels of indenture to reach the appropriate command. However, new interface devices have lessened this problem. Touch sensitive overlays and voice controls are two means to achieve easy communications. Boeing's new 777 uses touch sensitive cursor control devices so that the Captain and First Officer can achieve easier interaction with the multiple AMLCDs on the flight deck. Voice control is also becoming a viable means of pilot interaction with the cockpit. Some recent experiments with a connected word recognizer have shown that it is possible to use conversational commands and still achieve 99% recognition accuracy[48]. The ultimate goal of the conversational commands is to emulate the interaction of the

GR-1 and F-15E crews. An example of a conversational command is shown in Figure 10. The pilot has the capability with multiple paths to say four different phrases that mean the same thing. The pilot is not required to remember one specific phrase ("brittle" speech) to accomplish a task as in Figure 11. As long as there are alternate paths to obtain the same goal, conversational commands are possible.

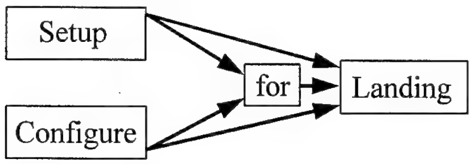


Figure 10. Multiple Path Voice Command

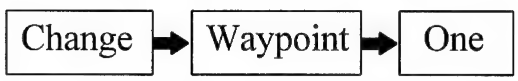


Figure 11. One path voice command

Guideline 3: Present Summarised Information

Even though an efficient means of communication exists between the pilot and the avionics system, it does not mean that information always flows in a clear and concise way between them. In modern military aircraft cockpits, pilots very often suffer from data overload, and with the inclusion of information from off-board sources, this overload problem may get worse. The designers of the displays can solve this problem by allowing the avionics system to present only summarised information. Icon based display formats, supplemented by text when necessary, are a very efficient way to achieve this goal. Steiner[49] presented pilots with system status information in both text form and through icons. For very simple displays the pilots performed equally well with either type of display; however, as the displays became more complex, as measured in bits, the icon based format was clearly superior. The pilots comments also supported the fact that icon based formats were easier to interpret and gave better situation awareness. As an example of additional work in this area, Way, Hobbs, Qualy-White, and Gilmour [50] developed a series of crew alerting system status displays. Figure 12 shows a summarisation of a fuel pump failure. In this display, the problem is depicted with icons, instructions for solving the problem are outlined, and the mission impact is stated.

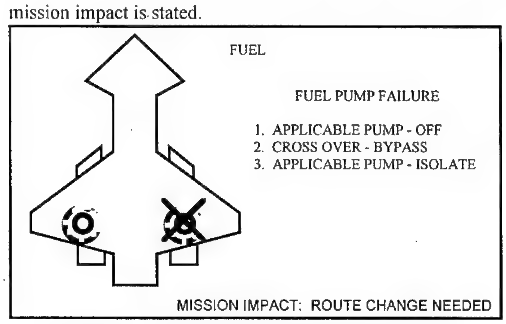


Figure 12. Crew Alerting System Status Display

14. CONCLUSION

Pilot aiding technology, coupled with adaptive interface controls and displays, have the potential of significantly helping pilots efficiently interact with the crewstation, while assisting in performing the mission more effectively. However, this technology by itself is no panacea; in fact, if not implemented in an intelligent manner, it could become a detriment to the pilot. The designer still needs to figure out how the subsystems and the pilot "play" together to present a clear picture of what the aircraft is doing, and to enable good decisions to be made about what needs to be changed, if required. The new technology is a two edged sword. It offers the designer virtually unlimited freedom to present information to the pilot; on the other hand, it also gives the designer the opportunity to swamp the pilot in data. The intelligent, cognitively compatible, glass cockpit will be the key to making the technology the pilots' friend, rather than their foe.

15. ACKNOWLEDGEMENTS

Dr M Haas and Dr G McMillan, USAF Research Laboratory, WPAFB., OH., US.

S Finnie, DERA Centre for Human Sciences.

Prof R Onken, University of the Armed Forces, Munich, Germany.

Dr B Brehmer and U Bergrund, FOA Department of Human Sciences, Sweden.

16. REFERENCES

1. Reising J. The Crew Adaptive Cockpit: Firefox, Here We Come. Proceedings of the 3rd Digital Avionics Conference, Fort Worth Texas. 1979, p204-207. CH1518-0/79/0000-0204. IEEE.

2. Emerson T.J., Reinecke M., Reising J.M. and Taylor R.M. The Human Electronic Crew: Can They Work Together? WRDC-TR-89-7008, Wright Patterson AFB, OH. 1989.

3. Emerson T.J., Reinecke M., Reising J. and Taylor R.M. The Human Electronic Crew: Is the Team Maturing? WRDC-TR-92-3078, Wright Patterson AFB, OH. July 1992.

4. R.M. Taylor and J. Reising, The Human Electronic Crew: Can We Trust the Team? WL-TR-96-3039, Wright Patterson AFB, OH., December 1995.

5. Taylor R.M., Reising J.M. and Onken R. (Eds), "The Human Electronic Crew: The Right Stuff?" Proceedings of the 4th International Workshop on Human Computer Teamwork, Kreuth, Germany, 23-26 September 1997 (In press).

6. Geddes N.D. and Shalin V.L. Intelligent Decision Aiding for Aviation. HFA Report. Swedish Centre for Human Factors in Aviation, Linkoping, Sweden. 8 September, 1997. ISSN 1402-7585. ISBN 91-7219-125-2

7. Geddes N.D. Associate systems: a framework for human-computer co-operation. In Smith M.J., Salvendy, G. and Koubek R.J. (Eds). Design of Human-Computer Systems: Social and Ergonomic Considerations. Advances in Human Factors and Ergonomics, 21B. 237-242. Amsterdam, Elsevier. 1997.

8. Egglestone R.G. Cognitive Interface Considerations for Intelligent Systems. AGARD-CP-520. April 1993. Paper 21, p1-16. NATO Advisory Group for Aerospace Research and Development, Neuilly sur Seine.

9. Miller C.A. and Riley V. Achieving the Associate Relationship: Lessons learnt from 10 years of Research and

Design. WL-TR-96-3039, p 155-161. Wright Patterson AFB, OH., December 1995.
10. Reising J.M. Must the Human-Electronic Crew Pass the Turing Test? WL-TR-96-3039, p 103-108. Wright Patterson AFB, OH., December 1995.
11. Egglestone R. Adaptive Interfaces as an Approach to Human-machine Co-operation. In, Smith M.J., Salvendy, G. and Koubek R.J. (Eds). Design of Human-Computer Systems: Social and Ergonomic Considerations. Advances in Human Factors and Ergonomics, 21B. 495-500. Amsterdam, Elsevier, 1997.
12. Sheridan T.B. and Verplanck W.L. Human and computer control of undersea teleoperators. Technical Report. Cambridge, MA: Massachusetts Institute of Technology, Man-Machine Laboratory. 1978.
13. Zachary W., Wherry R., Gleen, F., and Hopson, J. Decision situations, decision process and decision functions: Towards a theory-based framework for decision-aid design. Gathersberg, MD: National Bureau of Standards, Human Factors in \computer Systems Conference. 1982.
14. Taylor, R. M. and Selcon, S. J. "A Teamwork Model of Pilot Aiding: Psychological Principles for Mission Management System Design". AGARD-CP-504. March 1992, Paper 9. NATO Advisory Group for Aerospace Research and Development, Neuilly sur Seine.
15. Taylor, R. M. and Selcon, S. J. "Operator Capability Analysis: Picking the Right Team". AGARD-CP-520. April 1993. Paper 20, p1-17. NATO Advisory Group for Aerospace Research and Development, Neuilly sur Seine.
16. Taylor R.M, Shadrake R., and Haugh J. Trust and Adaptation Failure: An Experimental Study of Unco-operation awareness". WL-TR-96-3039, p87-92, Wright Patterson AFB, OH., December 1995.
17. Lucas A.T., Selcon S.J., Coxell A.W., Dudfield H.J., and O'Clarey N. Communication in the Human Electronic Crew: What can we learn from human teams? WL-TR-96-3039, Wright Patterson AFB, OH., December 1995.
18. Taylor R.M., Haugh J., and Shadrake R. Trust and Adaptation Failure: An Experimental Study of Unco-operation awareness". WL-TR-96-3039, Wright Patterson AFB, OH., December 1995.
19. Dreyfus H. and Dreyfus S.E. Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer. New York: The Free Press. 1988
20. Gerlach M. and Onken R. CASSY - The Electronic Part of a Human-Electronic Crew. WL-TR-96-3039, p140-145. Wright Patterson AFB, OH., December 1995.
21. Prevot T., Onken R., and Dudek H.-L. 'Knowledge-based Planning for Controlled Airspace Flight Operation as Part of a Cockpit Assistant'. AGARD-CP-504. March 1992, Paper 12. NATO Advisory Group for Aerospace Research and Development, Neuilly sur Seine.
22. Amalberti R and Deblon F. Cognitive modelling of fighter aircraft process control: A step towards an intelligent on board assistance system. International Journal of Man Machine Studies. 1992.
23. Champigneux G. and Joubert T. The Copilote Electronique. In Taylor R.M., Reising J.M. and Onken R. (Eds), "The Human Electronic Crew: The Right Stuff?" Proceedings of the 4th International Workshop on Human Computer Teamwork, Kreuth, Germany, 23-26 September 1997 (In press).
24. Lizza C.S., and Banks S.B. The Pilot's Associate: a co-operative knowledge based application, IEEE Expert, June 1991.
25. Lizza C.S., Rouse D.M., Small R. and Zenyuth J.P. Pilot's Associate: An Evolving Philosophy. WRDC-TR-92-3078, p149-154, Wright Patterson AFB, OH. July 1992.
26. Lynch R., Arbak C, Barnett B., and Olson J. Displays and Controls for the Pilot/Electronic Crewmember Team. WL-TR-96-3039, p 130-138, Wright Patterson AFB, OH., Dec 1995.
27. Krobusek R.D., Boys R.M., and Palko K.D. Levels of autonomy in a tactical electronic crewmember. WRDC-TR-89-7008, Wright Patterson AFB, OH. 1989.
28. Yadrick R., Judge C., and Riley V. Decision Support and Adaptive Aiding for Battlefield Interdiction Mission. WRDC-TR-92-3078, p122-126, Wright Patterson AFB, OH. July 1992.
29. Lynch R., Arback C.J., Barnett B., and Olson J.L. Displays and Controls for the Pilot/Electronic Crewmember Team. WL-TR-96-3039, p 130-134. Wright Patterson AFB, OH., December 1995.
30. Zenyuth J.P., Small R.L., Hammmer J.M., and Greenberg A.D. Principles for Interaction for Intelligent Systems. WL-TR-96-3039, p 162-166. Wright Patterson AFB, OH., December 1995.
31. Small R.L. Search Technologies Hazard Monitor: A Pilot Decision Aiding System. In Taylor R.M., Reising J.M. and Onken R. (Eds), "The Human Electronic Crew: The Right Stuff?" Proceedings of the 4th International Workshop on Human Computer Teamwork, Kreuth, Germany, 23-26 September 1997 (In press).
32. Joubert T., Salle S.E., Champigneux G., Grau J.V., Sassus P., and Le Doeuff. The Copilote Electronique Project: First Lessons as Exploratory Development Starts. WL-TR-96-3039, p 167-172. Wright Patterson AFB, OH., December 1995.
33. Ellis R.D., Hepworth R., Howells H., and Bickerton R.E. Structured and Analytical Knowledge Acquisition Methods for Tactical KBS Decision Aids. WL-TR-96-3039, p 59-64. Wright Patterson AFB, OH., December 1995.
34. Clare J., Peden C., and Smith R. Designing Real-time Decision Support for Future Systems and Scenarios. WL-TR-96-3039, p 65-70. Wright Patterson AFB, OH., December 1995.
35. Wittig T. and Onken R. Pilot Intent and Error Recognition as Part of a Knowledge Based Cockpit Assistant. AGARD-CP-520. April 1993. Paper 9, p1-10. NATO Advisory Group for Aerospace Research and Development, Neuilly sur Seine.
36. Miller C. and Goldman R. "Tasking" Interfaces: Associates that Know Who's the Boss. In Taylor R.M., Reising J.M. and Onken R. (Eds), "The Human Electronic Crew: The Right Stuff?" Proceedings of the 4th International Workshop on Human Computer Teamwork, Kreuth, Germany, 23-26 September 1997 (In press).
37. Taylor R.M. Human-Electronic Crew Teamwork. In, Smith M.J., Salvendy, G. and Koubek R.J. (Eds). Design of Human-Computer Systems: Social and Ergonomic Considerations. Advances in Human Factors and Ergonomics, 21B. Amsterdam, Elsevier. 247-250, 1997.
38. Taylor R.M. and Finnie S.E. The Cognitive Cockpit: Adaptation and Control of Complex Systems. In, Taylor R.M., Reising J.M. and Onken R. (Eds), "The Human Electronic

Crew: The Right Stuff?" Proceedings of the 4th International Workshop on Human Computer Teamwork, Kreuth, Germany, 23-26 September 1997 (In press).

39. Rassmussen J. Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering. New York: North Holland, 1986.

40. Hollnagel E. Context, Cognition and Control. In Y. Waern (Ed). "Co-operation in Process Management - Cognition and Information Technology". London: Taylor and Francis (In press).

41. Brehmer B. Dynamic Decision Making: Human Control of Complex Systems. Acta Psychologica, 81, North Holland,1992.

42. Hollnagel E. Control versus Dependence: Striking the Balance in Function Allocation. In, Smith M.J., Salvendy, G. and Koubek R.J. (Eds). Design of Human-Computer Systems: Social and Ergonomic Considerations. Advances in Human Factors and Ergonomics, 21B. 243-246. Amsterdam, Elsevier. 1997.

43. Powers, W.T. Behaviour-the Control of Perception. Chicago, Aldine, 1973.

44. Taylor M.M. Principles for Integrating Voice I/O in a Complex Interface. AGARD-CP-521. October 1992. Paper 21, p1-14. NATO Advisory Group for Aerospace Research and Development, Neuilly sur Seine.

45. Guevara K. Cognitive Architectures for Supporting Strategic Behaviours in Adaptive Systems. In, Taylor R.M., Reising J.M. and Onken R. (Eds), "The Human Electronic Crew: The Right Stuff?" Proceedings of the 4th International Workshop on Human Computer Teamwork, Kreuth, Germany, 23-26 September 1997 (In press).

46. Wilson and Rutherford. Mental models: theory and application in human factors. Human Factors, 13, 617-634. 1989.

47. Reising, J., Solz, T., Liggett, K., and Hartsock, D. A comparison of two head up display formats used to fly curved instrument approaches, In Proceedings of the 39th Annual Meeting of the Human Factors Society. Santa Monica: HFS. 1995.

48. Barry, T., Solz, T., Reising, J., and Williamson, D. The simultaneous use of three machine speech recognition systems to increase recognition accuracy, In *Proceedings of the National Aerospace and Electronics Conference*, (IEEE). 1994.

49. Steiner, B. and Camacho, M. Situation awareness: icons vs. alphanumerics. In Proceedings of the 33rd Annual Meeting of the Human Factors Society, Santa Monica: HFS. 1989.

50. Way, T., Hobbs, R., Qualy-White, J., and Gilmour, J. 3-D Imagery Cockpit Display Development, Technical Report WRDC-TR-90-7003, 1990.

A Conceptual Model For Understanding Computer-Augmented Distributed Team Communication and Decision Making

T. S. Andre, B. M. Kleiner, & R. C. Williges
Department of Industrial and Systems Engineering
Virginia Polytechnic Institute and State University
Blacksburg, VA USA 24061

SUMMARY

Distributed teams in dynamic environments provide both opportunities and technical challenges. Generally, distributed teams allow organizations to combine resources regardless of geographic dispersion to optimize organizational outcomes. However, the dispersion of a team introduces the technical challenge of providing effective communication between team members for appropriate decision making. The appropriate design and use of computer-augmented communication technology is essential for structuring teams for optimum performance. This paper presents a conceptual model for understanding the various input/output variables at work in distributed teams. The model is discussed in terms of appropriate research paradigms needed for validation, modeling, and development of communication metrics associated with effective communication and decision making.

1. BACKGROUND

Many modern and emerging U.S. Air Force information-based tasks such as mission planning and dynamic resource allocation involve teams which are often distributed and only linked electronically. Effective distributed team performance is highly dependent upon the appropriate design and use of computer-augmented communication technology to support team information management. A better understanding of the complex relationship among information tasks, team dynamics, and communication technology is acutely relevant to the U.S. Department of Defense. Recent research has again demonstrated the need for valid communication metrics (Ref 1, 2). Computer-augmented communication metrics are needed to evaluate the role of communication in distributed teams and will lead to a better understanding of how to improve distributed team performance in a variety of emerging U.S. Air Force complex information management systems. Kies, Williges, and Williges (Ref 3) point out that effective communication in these systems must simultaneously consider communication technology, conferencing technology, the physical environment, the social environment, and the task environment.

The emergence of distributed teams allows organizations to combine resources (e.g., information, expertise, production capacity, problem solving capability) regardless of geographic dispersion to optimize organizational outcomes. This is popularly called "virtuality". Virtuality in the organizational design literature typically assumes a relatively temporary, self-managed structure but empirical knowledge regarding optimal design is lacking. Cano and Kleiner (Ref 4) characterized distributed or virtual teams as having traditional physical infrastructure components, including geographical collocation and means for physical interactions, replaced by information and communication technology. Since virtual teams will play an important role in future defense scenarios, a better understanding of virtual team design and the factors which contribute to virtual team performance is needed.

Unfortunately, distributed team research is scarce and many empirical studies on team performance, in particular those relating to the effects of organizational and technological factors, have resulted in inconclusive results or contradictory conclusions. One possible explanation for the conflicting anecdotal and empirical evidence is the lack of a systematic understanding of the factors that affect team performance. Most anecdotal evidence lacks the depth of analysis necessary to truly identify the underlying determinants of team performance. In particular, there is little evidence on the effect of geographic dispersion, the main characteristic of dispersed or virtual teams. Until empirical evidence is available, it may be reasonable to assume that some of the same mechanisms that mediate collocated team performance operate in distributed teams.

While many authors use the terms "team" and "group" as synonyms, it is believed that an important team dimension is the extent to which the unit is a nominal group (i.e., in name only) or a cohesive team which shares a mental model. Communication metrics are important not only to measure and evaluate performance and behavior, but to improve performance as well. Accelerating the transformation process from group to team through new methods and metrics is quite important to performance-oriented organizations such as the U.S. Air Force. It is hypothesized then that communication effectiveness is an important requirement for effective decision making and determinant of a group's ability to mature into a cohesive team. Computer-augmented communication metrics are needed to evaluate the role of communication in group and

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

team decision making and will lead to a better understanding of how nominal groups evolve into effective team decision makers.

1.1 Team Decision Making

Decision making in the air combat environment requires the rapid processing of large amounts of data in a complex and unpredictable environment (Ref 5). This processing of data yields information for decision making, where decision making is actually comprised of a complex array of variables. Kleiner's (Ref 6) decision making framework considers the decision making system as three interrelated input/output subsystems which affect a fourth subsystem, the organizational subsystem. This framework defines a decision making subsystem (DMS), a decision tool subsystem (DTS), and a work process subsystem (WPS). The decision making subsystem is concerned primarily with decision makers converting information from decision tools into decisions for action in the mission. The process of converting information to a decision requires effective communication. The decision tool subsystem transforms data from the work process (mission) into information for decision making. In a team environment, data may be derived from several sources, necessitating an effective communication process. The focus of the WPS is the data-information-decision-action process, and therefore, criteria for the WPS are operationally defined in information flow terms. The DMS is a human-dominant subsystem; whereas, the DTS is predominantly a technical subsystem. Consequently, the overall mission or WPS is viewed as a sociotechnical subsystem. However, each can be described through analogous performance criteria which relate to critical components and interfaces of each subsystem. Subsystem-specific measures have been defined for each general performance criterion. The model uniquely considers the decision makers, their decision tools, and the mission in which their decisions are focused and will be expanded to focus more explicitly on the role of communication in team decision making.

Several theories exist to explain components of decision making within a team. The multi-level theory (MLT) proposed by Hollenbeck, Illgen, Sego, Hedlund, Major, and Phillips (Ref 7) is the most relevant to decision making in a distributed environment. Hollenbeck et al. developed MLT to test decision making performance for hierarchical teams with distributed expertise. This theory identifies three core team-level constructs in determining the relevant requirements for effective decision making accuracy. The three constructs include: (1) team informity -- degree to which the team is informed of all relevant cues associated with the decision; (2) staff validity -- degree to which lower level members help predict the true state of the decision; and (3) hierarchical sensitivity -- degree to which the team leader effectively weights staff judgments. Hollenbeck et al. used TIDE^2 (Team Interactive Decision Exercise for Teams Incorporating Distributed Expertise) as a laboratory task to represent multilevel theory team decision making in a command and control environment. They found that teams high on all three constructs performed significantly better than teams that were low on all three. Tower and Elliott (Ref 8) used TIDE^2 to demonstrate the positive role of communication efficiency as measured by wasted communication effort on teams with distributed expertise. They showed that team informity mediated both the relationship between communication effectiveness and team performance as well as the relationship between communication effectiveness and staff validity.

1.2 Team Effectiveness

Investigating human processes and performance at the team level is important for several reasons. First, there is a need for further empirical research on team effectiveness (Ref 9, 10). Second, different team types may represent different design and management requirements (Ref 11, 12). This implies the need to understand and define the determinants of effectiveness for different types of teams (Ref 13). General types of teams include *functional* teams which are homogeneous with respect to technical orientation of its members; *cross-functional* teams which combine members with different technical competencies; and *self-managing* teams which have autonomous control over such decisions as member selection, work design, and scheduling. While taxonomies exist for categorizing types of teams, few address the functions of teams or relate team functioning to performance. One notable exception is a framework proposed by Regian and Elliott (Ref 14) where teams are classified in terms of the degree to which the team is responsible for such classical functions as resource allocation, coordination, monitoring, information exchange, and adaptive problem solving. From these functions, metrics of team interdependencies and performance can be derived.

Most research models of team effectiveness are variations of McGrath's (Ref 15) model of task group effectiveness. Gladstein (Ref 16) has tested the task group effectiveness model by manipulating inputs and the group process in order to measure outcomes such as group performance and satisfaction. Inputs in the task group effectiveness model include individual skills and experience, group structure and composition, organizational resources and structure, and environmental factors. Most models of group performance consider group interaction processes as important factors influencing group performance. While member satisfaction is often the metric used in group performance measures it is a self-reported construct and is vulnerable to bias (Ref 17). Group interaction which is directly observable may be more reliable. Group interaction processes include metrics such as communication, adaptability, cooperation,

acceptance of suggestions or criticisms, giving suggestions or criticisms, group spirit, and coordination (Ref 18).

1.3 Decision Making Tools

1.3.1 *Communication support systems*

Kraemer and Pinsonneault (Ref 19) point out the importance of differentiating between two fundamentally different types of technology support systems. *Group Communication Support Systems*, or GCSS, serve primarily as information aids. Pinsonneault and Kraemer (Ref 20) identify the purpose of these systems as reducing communication barriers and supporting the communication process of the group. General examples of these systems are collaborative laboratories, electronic conference rooms, electronic chalkboard systems, etc. These systems provide information exchange and control and representational capabilities as described by Zachary (Ref 21) in his taxonomy of decision support techniques.

1.3.2 *Decision support systems*

Group Decision Support Systems, or GDSS, are designed to perform the more complex role of decision support. These systems utilize technology to support the decision making process by reducing the effect of our limited ability to process information (Ref 22). In this way, groups are capable of integrating the knowledge of all members into better decision making. GDSS supports the decision making process by providing structure for the decision making process in some way. Zachary (Ref 21) also describes some of the support techniques corresponding to GDSS. These functions include choice models, analysis and reasoning methods, judgment refinement techniques, and process models.

An extensive review of the study of decision support systems by DeSanctis and Gallupe (Ref 23) classified systems into three levels and identified some fundamental differences between systems at these different levels. Level 1 systems are equivalent to GCSS, and Level 2 systems equivalent to GDSS. They also identified Level 3 systems, which they defined as systems where the technology would actually intervene in the way group members interact by what they called "machine induced group communication patterns". To date, very little research exists in this area. Although Pinsonneault and Kraemer (Ref 20) combined Level 1 and Level 3 systems into the GCSS category, a more appropriate taxonomy may be to classify them separately, perhaps in a new category called Computer Mediated Communication Systems (CMCS) as suggested by Hiltz, Johnson, and Turoff (Ref 24).

Pinsonneault and Kraemer (Ref 20) point out that GDSS increase consensus, decease decision time, and increase satisfaction among group members; whereas, GCSS decreases cooperation and increases decision time among team members. Since GDSS supports decision making communications directly and GCSS provides general tools for communication, it can be hypothesized that the effect of communication on team performance is complex and may depend upon communication content categories and interactions with team tasks. Metrics which capture communication content are needed to test these hypotheses.

1.4 Team Communication

Communication is an essential element for most U.S. Air Force decision making tasks and requires effective integration for operators to perform the task and for observers to measure adequately the decision making process. A communication and decision making model for distributed teams is not complete unless there is some way to integrate the communication channels and provide the communication metrics that are sensitive to changes in the decision making process.

1.4.1 *Communication channels*

Kies, Williges, and Rosson (Ref 25) describe text, audio, and video as the three major communication channels used in computer-augmented communication. Each of these channels of communication has unique advantages and disadvantages, can affect communication independently, and can interact with the other two channels of communication depending upon the application task.

1.4.1.1 Text

Of the three communication channels, text represents the lowest bandwidth requirements. Hiltz, Johnson, and Turoff (Ref 26) and in a study replicated by Adrainson and Hjelmquist (Ref 27) found that text was better used for tasks in which the visual cues were not as critical for the particular task. Such tasks included the exchange of information, exchanging opinions, and generating ideas. It was demonstrated that tasks which relied on the transmission of gestures and facial expressions, such as persuasion, problem resolution, and getting acquainted with the other participants were better accomplished in a face-to-face session. Computer-augmentation of text-based tools can include brainstorming (Ref 28), nominal group technique (Ref 1) and the ability of group editors to enhance the writing quality when collaborators are not collocated as in distributed teams (Ref 29).

In addition to low bandwidth, text-based communication has the advantages of easy storage and preservation of communication history, ease of use, increased participation among team members, the availability of both synchronous and asynchronous formats, less conformity of opinions, and no dominate participant (Ref 30). On the other hand, text-based communication has the disadvantages of requiring more time to reach consensus and extreme

views are more likely to be presented (Ref 26, 27, 30, 31), difficulty in interpreting nuances of messages without verbal or visual cues (Ref 32), and coordination of communication among team members may become difficult (Ref 30, 33).

1.4.1.2 Audio

The audio channel plays a crucial role in human-human communication and is perhaps the most significant communication channel. Chapanis (Ref 34) in a series of studies found the auditory channel is the most critical to effective communication. Voice communication is more personal because it is often easily recognizable, and does not offer the de-personalizing effects of text (Ref 31). Short, Williams, and Christie (Ref 35) and Böcker and Mühlbach (Ref 36) concluded that an active auditory channel increases the sense of being located in the same physical location. Auditory communication, however, has the potential disadvantages of being transient, not easily storable for recall and manipulation, potentially overloading, and potentially resulting in poor localization (Ref 37).

1.4.1.3 Video

Video-based communication is becoming more feasible over computer networks due to faster computer processors, better compression algorithms, and faster networks. Kies, Williges, and Rosson (Ref 38) suggested that performance does not suffer under reduced frame rate and resolution of current technology video conferencing systems, but users' opinions of the degraded video quality are negative. Research has shown that conversation is mediated to a large extent with visual cues such as eye contact, facial expressions, gesture, and posture (Ref 39). Additional information such as participant interest, personality, and emotion is also conveyed through the visual channel. Other studies have focused on the effects of specific attributes of visual communication. For example, Harrison (Ref 40) discussed the role of the head and face in relaying demographic information, level of interest, and emotion. Argyle and Cook (Ref 39) discussed the importance of gaze in mediating conversation, and Mühlbach, Böcker, and Prussog (Ref 41) investigated technical solutions to increase the degree of mutual gaze.

Given the important role the visual channel plays in communication, it is curious that many studies show no task performance improvement when adding video to a communication system (Ref 42, 43, 44, 45, 46). Some researchers (Ref 47, 48) noted these studies have used subjects who are unfamiliar with each other, measured performance on a contrived task, and do not account for long-term usage. They sought to overcome these methodological limitations by taking a more ethnographic approach to studying the role of video in communication. While their findings suggest users prefer video and this medium affects the use of other communication modes, there remains no solid understanding of how video is used for real tasks over an extended period of time.

1.4.1.4 Communication channel comparison

Few studies have linked communication to decision making (Ref 6). Research on the utility of alternative communication media has been more common, but is a complex issue. Some studies have compared video to other media such as text and audio while others have focused on the differences between video and face-to-face communication. Studies generally reflect the objectives and backgrounds of the researchers, resulting in large differences between methodology, rigor, and interpretation. These differences make extracting commonalties from this large body of research difficult. Some researchers propose that the utility of the video channel is task-dependent (Ref 35, 44, 49). Interpersonal tasks, such as negotiation and persuasion can be assisted by the inclusion of a video channel, whereas information exchange tasks such as lectures and brainstorming are relatively immune to the effects of video (Ref 49). Williams (Ref 44) partitions tasks into cooperative and conflictful classes and argues that the former is, for the most part, insensitive to video while the latter benefits from a video channel.

1.4.2 *Observational procedures for measuring team communication*

Complex computer-augmented communication systems are best studied with a variety of complimentary measures, and a balance of unique methods is the most appropriate approach to studying these systems (Ref 50). For example, Anderson, Newlands, Mullin, Fleming, Doherty-Sneddon, and Van der Velden (Ref 51) studied video-mediated communication with task performance measures, subjective assessment, and direct conversation measures. Kies, Williges, and Rosson (Ref 25) describe ethnographic procedures that are needed to investigate communications in conferencing environments. Although the majority of the research on electronically-mediated video, audio, and text communication has been examined in highly controlled situations, observational methods are also useful in evaluating communications of distributed teams. Interviews, questionnaires, ethnomethodological observations, and conversation analysis measures provided a holistic understanding of video-mediated communication.

1.4.2.1 Verbal protocol analysis

The use of verbal reports as data can augment traditional data collection methods because they can describe strategy, engage operators, involve operators, distinguish between levels of operation, identify and promote learning, and explain individual differences (Ref 52). A "protocol" is a verbatim transcript of a decision-making or problem-

solving process (Ref 53). There are three types of verbalization: (1) reports of stimuli that remain constant and are available to the senses; (2) information retained in memory; and (3) information retained in long-term memory (Ref 54). Many studies rely on concurrent data which comprise verbal reports created while a subject performs a task. Retrospective protocols, which are generated after a task is performed and are believed to require long-term memory (LTM) retrieval, are needed to both validate concurrent protocols and to gain deeper insight.

Ericsson and Simon (Ref 54) have defended the use of verbal protocols as "hard" data. In addition to producing hard data, the use of protocols produces interesting data not otherwise obtained (Ref 55). Since there is much evidence for cognitive processes in judgment and discrimination (Ref 54), it follows that cognitive modeling of these processes can be realized through the use of protocols.

An alternative to protocol analysis is to use content analysis, where behavior is categorized, coded, and analyzed. For example, Sanquist and Fujita (Ref 56) were interested in design and evaluation issues. They predicted categories of protocols and subsequently evaluated the percentage of protocols falling into these categories. The authors indicated that such a categorization scheme could eliminate the use of costly prototypes. In another example of content-based approaches, the impact of active participation and communication of humans in automated systems was studied (Ref 57). In this study, seventeen categories of inter-operator communications were effectively used.

1.4.2.2 Interaction analysis

Audio-video tapes of team communications can be analyzed by an interaction analysis. Originally, the content of the team communications were classified into different types of communications such as declarative statements, questions, etc. by experts analyzing the audio communication recordings (Ref 58, 59). Recently, more sophisticated approaches such as the System of Multiple-Level Observation of Groups (SYMLOG) provides quantitative measures of group interactions (Ref 60). The SYMLOG measurement scheme has been used with fairly good reliability, but such techniques are tedious and require the presence of a trained expert (Ref 1, 2).

A more detailed content analysis consists primarily of recording the frequency of communication breakdown and communication turntaking among team members. For example, Kies (Ref 61) used an interaction analysis to isolate six classes of operationally-defined communication breakdown metrics: verbal turntaking breakdown, reference breakdown, visual breakdowns, audio breakdowns, and shared conversation breakdowns. Based upon the type of task involved, the specific metrics required for interaction analysis may vary.

1.4.2.3 Ethnomethodology

Hughes, Sharrock, Rodden, Kristoffersen, O'Brien, Rouncefield, and Calvey (Ref 62) discuss ethnomethodological data that were collected by reviewing the videotapes and looking for behavioral patterns, such as work flow organization, interaction with and use of artifacts, plans, and procedures, and mechanisms by which work activities were made aware to all group members. Metrics can be developed to isolate various communication themes central to distributed team performance. For example, Kies (61) used interviews, questionnaires, ethnomethodological observations, and conversation analysis measures to provide a holistic understanding of video-mediated communication. Six themes or behavioral patterns related to conversation fluidity, support tools, effects of time, work organization, group configuration, and visual issues emerged from these four analysis methods.

1.4.2.4 AT:ST ratio

The ratio between analysis time and sequence time is defined as the AT:ST ratio (Ref 63). Techniques based on real-time observation, such as the ones mentioned previously, have AT:ST ratios close to 1:1. As the variables of interest become more complex, the techniques used to measure them require increasing amounts of post-experimental processing and coding before the data is suitable for analysis. This results in increasing AT:ST ratios. For example, ratios between 3:1 and 10:1 have been reported for techniques aimed at analyzing video-based usability data (Ref 64). For the most complex types of analysis, such as the modeling of cognitive processes of users, ratios of 500:1 and even 5000:1 have been reported (Ref 65). Clearly, large AT:ST ratios impose a practical limitation on the ability to implement rich but complex communication measurement techniques (Ref 63).

Non-augmented observational procedures are prevalent in the literature. Observational procedures aimed at measuring outcome or perceived variables are the most common and simple to implement. Variables range from decision quality in a hypothetical survival case (Ref 66) to situational awareness in a flight simulation scenario (Ref 67). In many cases these variables can be measured by real-time observation, such as timing, counting, etc., or by post-experimental questionnaires. Many of these techniques require little or no post-experimental processing or coding of the data before analysis can begin.

1.4.3 *Computer-augmented procedures for measuring team communication*

Due to the complexities of using verbal reports as data, and the current trend to employ this technique for studying cognitively complex tasks, interest in automated and semi-automated knowledge acquisition tools and expert systems has increased. One area where advances in technology are clearly assisting in the measurement process of team communication is in the automated transcription of human speech. Once only possible through tedious and time consuming manual work, Automatic Speech Recognition (ASR) technology has finally become powerful enough for continuous speech, speaker independent voice recognition. Several system variables including recognition accuracy, recognition speed, recognition feedback, recognition errors, error correction, speaker mode, and vocabulary size can directly affect the acceptability and usability of ASR technology (Ref 68, 69, 70, 71).

Commercial applications of ASR requiring limited vocabularies are currently available for PC-compatible computers (Ref 72, 73). Some examples of commercially available packages which offer these capabilities are IBM Via Voice, Dragon Systems' Naturally Speaking, and Voice Recognition Systems' VRS package. All these systems implement ASR technology while only requiring off-the-shelf PC computer hardware platforms. In fact, the JABBER system (Ref 74) integrates speech recognition, lexicographic analysis, agenda management, temporal idiom recognition, and automatic annotations to analyze the interactions occurring in a meeting in real time and with almost no human intervention. Although the system is still in prototype stages, and it is designed for indexing meeting recordings for latter retrieval; its capabilities indicate the current directions in automated communication metrics.

The next step in augmenting the measurement of team communications is the coding of speech utterances into a format suitable for analysis. Interest on these technologies has also being on the rise. For example, Kleiner and Drury (Ref 52) described the use of the Ethnograph (Ref 75) for content analysis of skill-based tasks. Boose (Ref 76) reviewed the Expertise Transfer System, a computer program that interviews experts and helps them build expert systems. Sanderson, James, and Sieder (Ref 77) reviewed SHAPA, an interactive verbal protocol tool. The MacSHAPA system (Ref 78), an extension of the original SHAPA, incorporates text, audio, and video-based content analysis.

The advantage of these and other similar systems is that they provide a systematic means for exploring cognitive processes. Also, by automating or increasing the processing time and coding of raw textual, verbal, and visual data, these techniques help reduce the AT:ST ratio for the most complex measurement techniques, making them more feasible to implement. It is important to note that although these systems greatly assist in the data collection process, they are by no means automated at their current stage and expert input is almost always required to complement the computer analysis.

As noted by Sanderson et al. (Ref 78), there are potentially a number of computer-supported approaches that can be taken, and investigators will often tailor the approach on the basis of the goals and constraints associated with a particular study.

In summary, most of the observational metrics described in this paper require expert observers who have been trained to use these procedures. In addition, these techniques usually require considerable time to collect and summarize the results thereby increasing the overall cost and reliability of these metrics. Since no fully automated approaches have been identified, there remains a need to automate these methods and metrics. Effective automation of these methods and metrics depends on a system level understanding of the team communication and decision making components where outputs from one subsystem provide the inputs to the next subsystem in an iterative fashion. A conceptual model of team communication and decision making can provide such an understanding.

2. CONCEPTUAL MODEL OF TEAM COMMUNICATION AND DECISION MAKING

The conceptual model shown in Figure 1 is based on three interdependent input/output subsystems: (1) *decision making team*; (2) *work process*; and (3) *decision making tools*. The model uniquely considers the decision makers, their decision tools and the mission on which their decisions are focused. The model currently focuses on a decision maker or group of decision makers managing a work process, team, or organization. The iterative aspect of the model is behavior oriented, focusing on team variables such as decisions, actions, measures, data, information portrayal, and information perception. A task dimension is provided by the linear flow of the work process, where social and technical variables influence the work and information flow. Communication is achieved through the "pipeline", where text, audio, and video are viewed as interacting modalities to accomplish both the behavior and task elements of the system. As groups evolve into teams, this model provides the necessary dependent variables to measure the qualitative and quantitative differences between the two. During experimentation, it is expected to see a measurable difference in output (e.g., performance, satisfaction, etc.) for teams compared to groups.

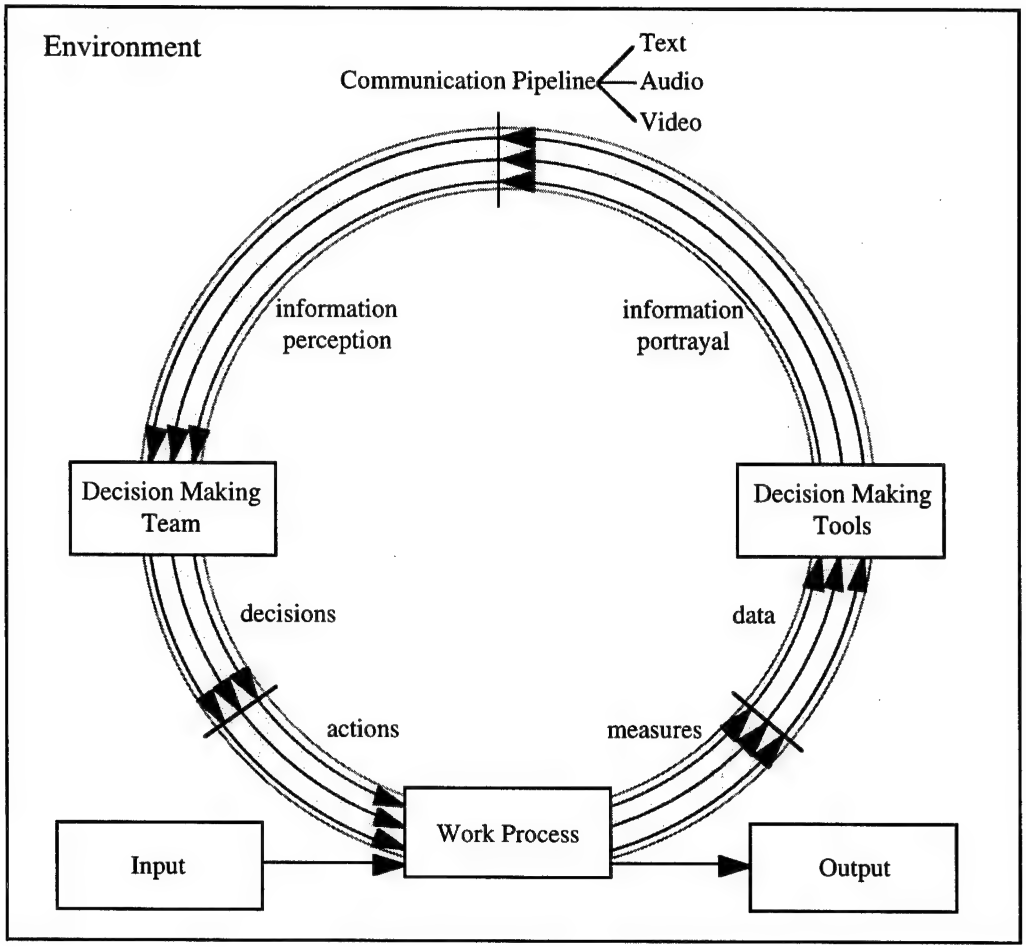


Figure 1. Conceptual model of team communication and decision making in a distributed environment.

Several different theories and models were integrated in constructing the model to provide a complete picture of the independent variables that influence distributed team communication and decision making. The independence of different theories and models make them sufficient for specific tasks and environments, but insufficient in terms of addressing other relevant aspects of the entire decision making system (e.g., tools, communication pipeline, work process). Kleiner (Ref 6) discusses the decision making, decision tool, and work process components within a management system, but does not specifically address teams or describe the communication channels needed for decision making. In the Hollenbeck et al. (Ref 7) model, the focus is on decision making accuracy related to team-level constructs such as team informity, staff validity, and hierarchical sensitivity. This model is insufficient in terms of addressing decision making tools and the communication modalities. Gladstein's (Ref 16) model focused on task group effectiveness and the group process, but did not address the decision making tools or communication pipeline for accomplishing the tasks. Gladstein concluded that theories of group effectiveness need to be revised to include the way in which teams manage interactions across their boundary and the impact of the organizational context.

Kleiner's (Ref 6) framework provides the analytical map for each of the subsystems in the model. With Kleiner's framework providing the map, the other models and theories fill in the details for looking at decision making in a distributed environment. Kies, Williges, and Williges (Ref 3) and Kies, Williges, and Rosson (Ref 25) provide the necessary components of communication to complete the model. Text, audio, and video combine to provide the necessary technology to transfer successfully information

from the decision making tools to the decision making team. Once the communication pipeline is established between the subsystems, Kleiner's framework provides the relevant dependent measures related to these subsystems.

Each subsystem in the model can be characterized by its own dependent variables. To illustrate this, Figure 2 shows a view of the decision making team where information is provided by the decision making tools and converted to decisions and actions for the work process.

Through this subsystem, several dependent variables become relevant. Some examples include tool quality, information quality, decision process quality, and quality of results. Each of the dependent variables can be operationalized and measured using subjective, objective data, or both. As an example, the dependent measure of information quality can be operationalized with the question "Is information what is expected, needed, or wanted?" This question can be answered subjectively by team personnel on a Likert-type survey instrument or objectively by deriving frequency counts when team members use information for an effective result.

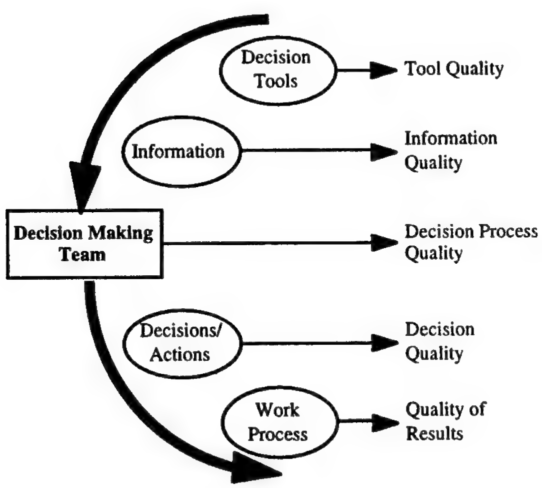


Figure 2. Quality dependent variables for the decision making team subsystem.

Figure 3 illustrates some of the dependent variables for the work process subsystem. Here, decisions and actions are used to accomplish the work and provide data for further transformation using decision making tools.

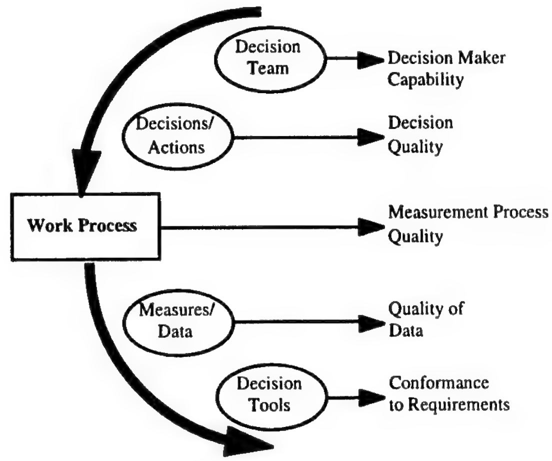


Figure 3. Quality dependent variables for the work process subsystem.

Figure 4 illustrates some of the dependent variables for the decision making tool subsystem. The focus in the decision making tool subsystem is the transformation of raw data from the work process to usable information for the decision making team.

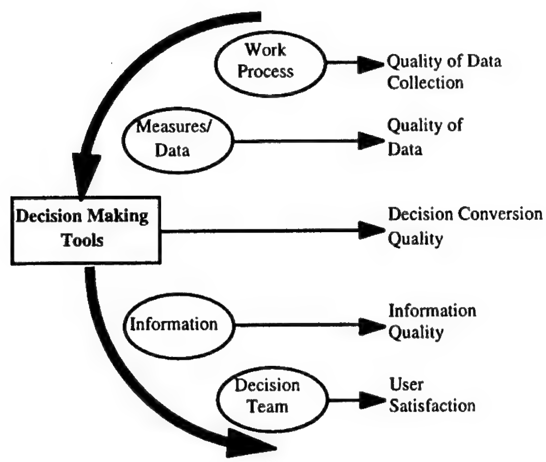


Figure 4. Quality dependent variables for the decision making tools subsystem.

3. RESEARCH FOCUS

This conceptual model of team communication, when combined with computer-augmented communication metrics, is expected to lead to a better understanding of team decision making and performance. From a theoretical standpoint, new knowledge will be gained about the interaction of communication processes with team functioning (Ref 14) and the role of communication in the various components of decision making performance (Ref 6). Three research tasks are relevant in order to develop and validate a suite of computer-augmented metrics for evaluating both the process and content of team communications: (1) task environment analysis; (2) metric development; and (3) metric validation as shown in Figure 5.

3.1 Task Environment Analysis

Identification of relevant tasks serve as inputs to the model of team communication, and the development of automated metrics play an important role in understanding team variables. Several subtasks are relevant to this analysis as illustrated in Table 1.

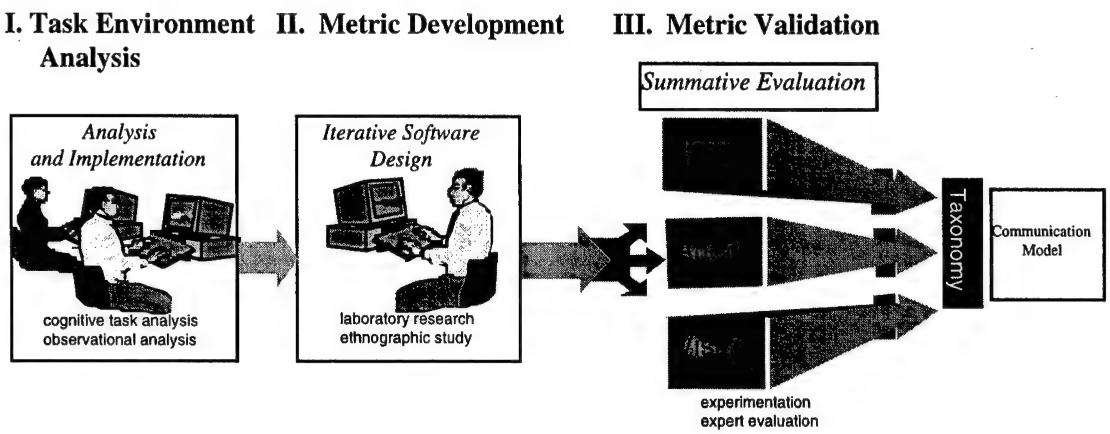


Figure 5. Research process flow for developing the communication model.

TABLE 1. Relevant subtasks for analyzing the task environment

Subtask	Focus
1. Perform cognitive task analysis	Establish relevant communication threads of team communication content.
2. Perform observational analysis	Defining communication threads in the team task that are amenable to computer-augmented capture and metric generation.
3. Collect concurrent and retrospective verbal protocols	Protocols help refine the cognitive task analysis. Performance with and without verbalization can be compared.
4. Embed metrics	Answer basic questions about team performance using a small set of metrics within the task.

3.2 Metric Development

The objective of metric development is to develop a suite of communication metrics which can be used in the task environment. The general approach to follow in developing these communication metrics is an *iterative design* procedure of software interface design emphasizing laboratory research and formative evaluation (Ref 25). Iterative design involves specifying and developing a prototype of the computer-augmented communication metric which is then re-designed in an iterative fashion using formative evaluation techniques until the metric reaches a desired level of reliability and usability. Specific subtasks to address in this iterative metric development process are summarized in Table 2.

3.3 Metric Validation

Metric validation primarily involves *summative evaluation* conducted in a laboratory setting using the task environment with its embedded communication metrics. Four major subtasks of validation are relevant as identified in Table 3.

TABLE 2. Specific subtasks to address in the iterative metric development process

Subtask	**Focus**
1. Implement PC-compatible ASR	Develop metrics of audio speech among team members using speech recognizers.
2. Perform ethnographic research	Results are used to develop and refine prototype computer-augmented communication metrics.
3. Compare direct observations with the cognitive task analysis	Find consensus communication threads and key word themes present in team communications.
4. Perform word and communication thread spotting	Techniques help generate metrics for textual and auditory team communications.
5. Specify candidate automated metrics	Address the content as well as process of team communication (e.g., frequency of communications, team member designations, number of exchanges, etc.).

TABLE 3. Specific subtasks for metric validation

Subtask	**Focus**
1. Perform expert evaluation	Compare results of automated procedures to direct observations by experts.
2. Analyze the AT:ST ratio	Demonstrate reduction in the AT:ST ratio
3. Conduct team communication experiments	Tie back to theory through efficient sequential experimental procedures using fractional factorial designs and central-composite design (Ref 79, 80, 81).
4. Develop communication metric taxonomy and guidelines	Taxonomy should provide team communication design guidelines for structuring team tasks.

The research tasks and subtasks are designed to help discover the underlying model for a particular application environment and the relevant communication metrics. These metrics can also be used to generate team communication design guidelines for structuring teams tasks. Some of the relevant research questions to address when developing design guidelines are identified in Table 4.

CONCLUSION

Studying team decision making requires the guidance of theory and computer-augmented metrics in order to understand fully the relevant variables that contribute to effective team performance. Traditional measures of team communication usually require direct observations by trained observers, have questionable reliability and validity, do not address communication content adequately, and are both time consuming and costly. Automated computer-embedded metrics of team communication can alleviate these shortcomings and help address both the process and content of team communication. Research is needed to test various models when applied to specific task environments. The conceptual model of team communication and decision making presented in this paper can guide the metric development. As metrics of team decision making and communication develop, researchers and practitioners will gain a better understanding of how to form and command teams for optimum performance in specific tasks.

TABLE 4. Research questions for developing team design guidelines

Subsystem	Question
Communication	• How do textual and auditory team communications compare in terms of communication efficiency and team performance? • Can various forms of structured team communications facilitate team performance? • How do communication processes interact with team functions? • How do communication modalities interact with different team functions and tasks? • Do various components of communication content interact with team functions?
Metrics	• To what extent can communication metrics and respective feedback be used to go beyond monitoring and evaluating performance and improve performance? • How do communication metrics relate to the various components of decision making performance? • How do metrics of team content relate to team performance?
Decision Making	• What is the relationship between communication and decision making for different tasks by different players in task management? • Precisely when and under what conditions does a nominal group become a cohesive team and what role does communication or DSS play?
Work Process	• What is the role of command evaluation and feedback in team performance? • How do communication processes and metrics differ as a function of type of task?
Tools	• How do decision support systems and communication support systems make the communication pipeline more effective?

REFERENCES

1. Hacker, M.E., "The Effect of Decision Aids on Work Group Performance", Unpublished Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1997.

2. Meredith, J.W., "Empirical Investigation of Sociotechnical Issues in Engineering Design", Unpublished Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1997.

3. Kies, J.K., Williges, R.C., and Williges, B.H., "Desktop Video Conferencing: A Systems Approach" in M. Helander (Ed.), "Handbook of Human-Computer Interaction (2nd Edition)", Amsterdam: Elsevier Science (North-Holland), 1997, pp. 979-1002.

4. Cano, A. and Kleiner, B.M., "Sociotechnical Design Factors of Virtual Team Performance", Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting, Philadelphia, PA, September 1996, pp. 786-790

5. Endsley, M.R. and Smith, R.P., "Attention Distribution and Decision Making in Tactical Air Combat", Human Factors, *38*, 2, 1997, pp. 232-249.

6. Kleiner, B.M., An Integrative Framework for Measuring and Evaluating Information Management Performance", International Journal of Computers and Industrial Engineering, *32*, 3, 1997, pp. 545-555.

7. Hollenbeck, J.R., Illgen, D.R., Sego, D., Hedlund, J., Major, D., and Phillips, J., "Multi-Level Theory of Team Decision Making: Decision Makingin Teams Incorporating Distributed Expertise", Journal of Applied Psychology, *80*, 2, 1995, pp. 292-317.

8. Tower, S.L. and Elliott, L.R., The Role of Communication Efficiency in Teams With Distributed Expertise: Application of the Multi-Level Theory", Proceedings of the International Association of Management Conference, Toronto, Canada, 1996

9. Goodman, P.S., Ravlin, E., and Schminke, M., "Understanding Groups in Organizations", in L.L. Cummings and B.M. Staw (Eds.), "Research in Organizational Behavior", Greenwich, CT: JAI Press, Inc., 1987.

10. McGrath, J.E., "Studying Groups at Work: Ten Critical Needs for Theory and Practice", in P.S. Goodman (Ed.), "Designing Effective Work Groups", San Francisco: Jossey-Bass Publishers, 1986.

11. Mohrman, S.A., "Integrating Roles and Structure in the Lateral Organization", in J.R. Galbraith and E.E. Lawler III (Eds.), "Organizing for the Future: The New Logic for Managing Complex Organizations", San Francisco: Jossey-Bass Publishers, 1993.

12. Parker, G.M., "Cross-Functional Teams: Working With Allies, Enemies, and Other Strangers", San Francisco: Jossey-Bass Publishers, 1994.

13. Goodman, P.S., "The Impact of Task and Technology on Group Performance", in P.S. Goodman (Ed.), "Designing Effective Work Groups", San Francisco: Jossey-Bass Publishers, 1986.

14. Regian, J.W. and Elliott, L.R., "Intelligent Tutoring for Team Performance: Theory-based Development of an Enhanced Platform for Team Training and Performance Research", Brooks AFB, TX: Air Force Armstrong Laboratory, 1997.

15. McGrath, J. E., "Social Psychology: A Brief Introduction", New York: Holt, Rinehart, and Winston, 1964.

16. Gladstein, D., "Groups in Context: A Model of Task Group Effectiveness", Administrative Science Quarterly, *29*, 1984, pp. 499-517.

17. Muckler, F. and Seven, S., "Selecting Performance Measures: Objective Versus Subjective Measurement", Human Factors, *34*, 4, 1992, pp. 441-455.

18. McIntyre, R. and Salas, E., "Measuring and Managing for Team Performance: Emerging Principles From Complex Environments", in R. A. Guzzo and E. Salas (Eds.), "Team Effectiveness and Decision Making in Organizations", San Francisco: Jossey-Bass Publishers, 1995, pp. 9-45.

19. Kraemer, K.L. and Pinsonneault, A., "The Implications of Group Support Technologies: An Evaluation of the Empirical Research", Proceedings of the Hawaii International Conference on System Science, 1989, pp. 326-336.

20. Pinsonneault, A. and Kraemer, K.L., "The Effect of Electronic Meetings on Group Processes and Outcomes: An Assessment of the Empirical Research", European Journal of Operational Research, *46*, 1990, pp. 143-161.

21. Zachary, W., "A Cognitively Based Functional Taxonomy of Decision Support Techniques", Human–Computer Interaction, 1986, *2*, pp. 25-63

22. Saaty, T.L., "Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World", Pittsburgh: RWS Publications, 1990.

23. DeSanctis, G. and Gallupe, R.B., "A Foundation for the Study of Group Decision Support Systems", Management Science, 1985, *33*, 5, pp. 589-609.

24. Hiltz, R.S., Turoff, M., and Johnson, K., "Experiment in Group Decision Making: Disinihbition, Deindividualization, and Group Process in Pen Name and Real Name Computer Conferences", Decision Support Systems, 1988, *5*, pp. 217-232.

25. Kies, J.K., Williges, R.C., Rosson, M.B., "Studying Computer-Supported Cooperative Work: A Review of Research Issues and Strategies", Journal of the American Society for Information Science, in press, to appear.

26. Hiltz, S.R., Johnson, K., and Turoff, M., "Experiments in Group Decision Making: Communication Process and Outcome in Face-To-Face Versus Computerized Conferences", Human Communication Research, 1987, *13*, 2, pp. 225-252.

27. Adrianson, L. and Hjelmquist, E., "Group Process in Face-To-Face and Computer Mediated Communication", Behaviour and Information Technology, 1991, *10*, 4, pp. 281-296.

28. Stefik, M., Bobrow, D.G., Foster, G., Lanning, S., and Tatar, D., "WYSIWYS Revised: Early Experiences with Multiuser Interfaces", ACM Transactions on Office Information Systems, 1987, *5*, 2, pp. 147-167.

29. Green, C. and Williges, R.C., "Evaluation of Alternative Media Used with a Groupware Editor in a Simulated Telecommunications Environment", Human Factors, 1995, *37*, 2, pp. 283-289.

30. Kiesler, S., Siegel, J., and McQuire, T.W., "Social Psychological Aspects of Computer-Mediated Communication", American Psychologist, 1984, *39*, 10, pp. 1123-1134.

31. Sproull, L. and Kiesler, S., "Connections: New Ways of Working in the Networked Organization", Cambridge, MA: MIT Press, 1991.

32. Garter, L.E. and Wellman, B., "Social Impacts of Electronic Mail in Organizations: A Review of the Research Literature", Technical Report, Center for Urban and Community Studies and the Ontario Telepresence Project, Toronto, Canada: University of Toronto, 1993.

33. Ellis, C.A., Gibbs, S.J., and Rein, G.L., "Groupware: Some Issues and Experiences", Communications of the ACM, 1991, *34*, 1, pp. 39-58.

34. Chapanis, A., "Interactive Human Communication", Scientific American, 1975, *232*, 3, pp. 36-42.

35. Short, J., Williams, E., and Christie, B., "The Social Psychology of Telecommunications", London: Wiley, 1976.

36. Böcker, M. and Mühlbach, L., "Communicative Presence in Videocommunications", Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, Seattle, Washington, September, 1993, pp. 249-253.

37. Gaver, W.W., "The Affordances of Media Spaces For Collaboration", Proceedings of CSCW '92: Computer Supported Cooperative Work., New York, NY: The Association for Computing Machinery, 1992.

38. Kies, J.K., Williges, R.C., and Rosson, M.B., "Evaluating Desktop Video Conferencing for Distance Learning Applications", Computers and Education, 1997, *28*, 2, pp. 79-81.

39. Argyle, M. and Cook, M., "Gaze and Mutual Gaze", Cambridge, U. K.: Cambridge University Press, 1976.

40. Harrison, R. P., "Beyond Words: An Introduction to Nonverbal Communication", Englewood Cliffs, NJ: Prentice-Hall, Inc., 1974.

41. Mühlbach, L., Böcker, M, and Prussog, A., "Telepresence in Videocommunications: A Study on Stereoscopy and Individual Eye Contact", Human Factors, 1995, *37*, 2, pp. 290-305.

42. Ochsman, R.B. and Chapanis, A., "The Effects of 10 Communication Modes on The Behavior of Teams During Co-Operative Problem-Solving", International Journal of Man-Machine Studies, 1974, *6*, pp. 579-619.

43. Weeks, G.D. and Chapanis, A., "Cooperative Versus Conflictive Problem Solving in Three Telecommunication Modes", Perceptual and Motor Skills, 1976, *42*, pp. 879-917.

44. Williams, E., "Experimental Comparisons of Face-To-Face and Mediated Communication: A Review", Psychological Bulletin, 1977, *84*, 5, pp. 963-976.

45. Gale, S., "Adding Audio and Video to an Office Environment", in J.M. Bowers and S.D. Benford (Eds.), "Studies in Computer Supported Cooperative Work", New York, NY: Elsevier Science, 1991.

46. Masoodian, M, Apperley, M., and Fredrickson, L., "Video Support for Shared Work-Space Interaction: An Empirical Study", Interacting with Computers, 1995, *7*, 3, pp. 237-253.

47. Tang, J.C. and Isaacs, E., "Why Do Users Like Video? Studies of Multimedia-Supported Collaboration", Computer Supported Cooperative Work (CSCW), 1993, *1*, pp. 163-193.

48. Isaacs, E.A. and Tang, J.C., "What Video Can and Cannot Do for Collaboration: A Case Study", Multimedia Systems, 1994, *2*, pp. 63-73.

49. Angiolillo, J.S., Blanchard, H.E., and Israelski, E.W., "Video Telephony", AT&T Technical Journal, May/June, 1993, pp. 7-20.

50. Watts, L., Monk, A., and Daly-Jones, O., "Inter-Personal Anwareness and Synchronization: Assessing the Value of Communication Technologies", International Journal of Human-Computer Studies, 1996, *44*, pp. 849-873.

51. Anderson, A.H., Newlands, A., Mullin, J., Fleming, A.M., Doherty-Sneddon, G., and Van der Velden, J., "Impact of Video-Mediated Communication on Simulated Service Encounters", Interacting with Computers, 1996, *8*, 2, pp. 193-206.

52. Kleiner, B.M. and Drury, C.G., "The Use of Verbal Protocols to Understand and Design Skill-Based Tasks", International Journal of Human Factors in Manufacturing, in press, to appear.

53. Glass, A.L., Holyoak, K.J., and Santa, J.L., "Cognition", Reading, Massachusetts: Addison-Wesley, 1979.

54. Ericsson, K.A. and Simon, H.A., "Protocol Analysis: Using Verbal Reports as Data. Cambridge", Massachusetts: The MIT Press, 1984.

55. Bainbridge, L., "Analysis of Verbal Protocols From a Process Control Task", in Edwards and Lees (Eds.), "The Operator in Process Control", London: Taylor and Francis, 1974.

56. Sanquist, T.F. and Fujita, Y., "Protocol Analysis and Action Classification in The Education of an Advanced Annunciator System Design", Proceedings of the Human Factors Society 33rd Annual Meeting, Denver, Colorado, October 1989, pp. 1064-1067.

57. Idaszak, J.R., "Human Operators in Automated Systems: The Impact of Active Participation and Communication", Proceedings of the Human Factors Society 33rd Annual Meeting, Denver, Colorado, October, 1989, pp. 778-782.

58. Bales, R.F., "Interaction Process Analysis: A Method for the Study of Small Groups", Cambridge, MA: Addison-Wesley, 1950.

59. Williges, R.C., Johnston, W.A., and Briggs, G.E., "Role of Verbal Communication in Teamwork", Journal of Applied Psychology, 1966, *50*, pp. 473-478.

60. Polley, R., Hare, A., and Stone, P., "The SYMLOG Practitioner: Applications Small Group Research", New York: Prager, 1988.

61. Kies, J.K., Empirical Methods for Evaluating Video-Mediated Collaborative Work", Unpublished Doctoral Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1997.

62. Hughes, J.A., Sharrock, W., Rodden, T., Kristoffersen, S., O'Brien, J., Rouncefield, M., and Calvey, D., "An Ethnography Handbook", in CSCW Requirements Development, COMIC Technical Report D 2.4, Lancaster, UK: Lancaster University, 1995.

63. Sanderson, P.M. and Fisher, C., "Exploratory Sequential Data Analysis: Foundations", Human-Computer Interaction, 1994, *9*, pp. 251-317.

64. Nielsen, J., Usability Engineering, San Diego: Academic Press, 1993.

65. Ritter, F.E. and Larkin, J.H., "Developing Process Models as Summaries of HCI Action Sequences", Human-Computer Interaction, 1994, *9*, pp.345-383.

66. Brannick, M.T., Prince, A., Prince, N., and Salas, E., "The Measurement of Team Processes", Human Factors, 1995, *37*, 3, pp. 641-651

67. Turoff, M. and Hiltz, S.R., "Computer Support for Group Versus Individual Decisions", IEEE Transactions on Communication, 1982, COM-30(1), pp. 82-90.

68. Atal, B.S., "Speech Technology in 2001", in D.B. Roe and J.G. Wilpon (Eds.), "Voice Communication Between Humans and Machines", Washington, D.C.: National Academy Press, 1994, pp. 467-481.

69. Baber, C., "Speech Technology in Control Room Systems: A Human Factors Perspective", New York: Ellis Horwood, 1991.

70. Baber, C., Usher, D.M., Stammers, R.B., and Taylor, R.G., 1992, "Feedback Requirements for Automatic Speech Recognition in The Process Control Room", International Journal of Man-Machine Studies, *37,* pp. 703-719.

71. Kamm, C., "User Interfaces for Voice Applications", in D.B. Roe and J.G. Wilpon (Eds.), "Voice Communication Between Humans and Machines", Washington, D.C.: National Academy Press, 1994, pp. 422-442.

72. Baber, C. and Noyes, J., "Automatic Speech Recognition in Adverse Environments", Human Factors, 1996, *38*, 1, pp. 142-156.

73. Rudnicky, A.I., Hauptmann, A.G., and Lee, K.F., "Survey of Current Speech Technology", Communications of the ACM, 1994, *37*, 3, pp. 52-58.

74. Kazman, R., Al-Halimi, R., Hunt, W., and Mantei, M., "Four Paradigms for Indexing Video Conferences", IEEE Transactions on Multimedia, Spring, 1996, pp. 63-73.

75. Seidel, J.V., Kjolseth, R., and Seymour, E., "The Ethnograph", Littleton, Colorado: Qualis Research Associates, 1988.

76. Boose, J.H., "Expertise Transfer for Expert System Design", Amsterdam: Elsevier, 1986.

77. Sanderson, P.M., James, J.M., and Sieder, K.S., "SHAPA: An Interactive Software Environment for Protocol Analysis", Ergonomics, 1989, *32*, 11, pp. 1271-1302.

78. Sanderson, P.M., Scott, J.J.P., Mainzer, J., Johnston, T., Watanabe, L.M., and James, J.M., "MacSHAPA and the Enterprise of Exploratory Sequential Data Analysis (ESDA)", International Journal of Human-Computer Studies, 1994.

79. Williges, R.C., "Review of Experimental Design", in J. Weimer (Ed.), "Research Techniques in Human Engineering", Englewood Cliffs, NJ: Prentice Hall, 1995, pp. 49-71.

80. Williges, R.C. and Williges, B.H., "Integrated Research Paradigm for Complex Experimentation", Proceedings of the Human Factors Society 33rd Annual Meeting, Santa Monica, CA, 1989, pp. 606-610

81. Williges, R.C., Williges, B.H., and Han, S.H., "Sequential Experimentation in Human-Computer Interface Design", in R. Hartson and D. Hix (Eds.), "Volume IV, Advances In Human-Computer Interaction", New York: Ablex, 1993, pp. 1-30.

Modeling the Decisionmaking Process and Performance of Airborne Warning and Control System (AWACS) Weapons Directors.

Linda R. Elliott
Samuel G. Schiflett
AFRL/HEAB Bldg. 170
Training Effectiveness Research Division
Human Effectiveness Directorate
Air Force Research Laboratory
2504 Gillingham Drive
Brooks AFB, TX. 78235-0482
U.S.A.

Mathieu A. Dalrymple
Human Systems Technology
Veridian, Inc.
2504 Gillingham Drive, Suite 201
San Antonio, TX. 78235-0482
U.S.A.

SUMMARY

In this paper we describe our investigation of team effectiveness in a dynamic, high workload environment, using our Command, Control, and Communications Simulation, Training And Research System (C^3STARS) as a microcosm of Command and Control Warfare. The primary research goal was to investigate relationships among measures of team effectiveness and performance and ascertain the effect of ambiguous communications on individual and team decisionmaking. Airborne Warning and Control System (AWACS) weapons directors (WDs) participated as subjects within three-person teams, in highly interdependent roles. Each team performed in six 3-hour air campaign scenarios. The scenarios were systematically developed to be equivalent in workload, each presenting participants with four 20-minute "waves" of high workload, as manipulated by time pressure, complexity, and ambiguity. Each wave contained 6 embedded decision events, targeted across each of the three team members. For each decision event, information was communicated to one team member which should be handed off to another team member. WD communications were captured and assessed for predicted relationships with performance and mission accomplishment. Ambiguity was manipulated by presenting information of varying levels of certainty. In addition, various measures of outcome measures of mission effectiveness are described, discussed, and related to measures of communication effectiveness. Results demonstrate qualitative differences in various measures of mission effectiveness. Measures of decisionmaking effectiveness were predictive of overall mission outcomes. Issues with regard to measurement of mission and team effectiveness are discussed.

1. INTRODUCTION

"Winning the information war is the new challenge for military commanders. Having highly accurate intelligence on the composition, location and movement of enemy forces allows commanders to employ their own forces to achieve a decisive advantage by delivering either a knock-out blow or protecting their own limited forces from attack".

This quote from the Global Defence Review (Ref 1) underlines the fundamental role of technology and communications for battlespace superiority. A core component of battlespace command, control, and communications (C^3) environment is the Airborne Warning and Control System (AWACS) command post. "The E-3 Sentry is the most sophisticated airborne command post ever devised. The airframe chosen for the job was the Boeing Model 707, which was equipped with state-of-the art electronics inside and topped with a 30-foot rotating radome" (Ref 2).

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

In the 1991 Gulf War, a network of USAF AWACS controlled more than 1,000 aircraft sorties a day throughout the middle east. Closely linked with AWACS radar systems are electronic signals and intelligence gathering aircraft that provide mutually supporting information to commanders. Because of its large airframe and fuel capacity the AWACS has extreme endurance, making it possible to remain on station over remote regions for long periods. Its multiplicity of communications links means it can talk to a large number of other aircraft, ships or ground stations. It is clear these capabilities challenge human-systems performance—to coordinate complex and dynamic information from multiple sources, in a high-stress, high-stakes environment, over a sustained period of time.

AWACS Weapons Directors (WDs) coordinate communications received from numerous others, such as other WDs, base operations, and friendly pilots. To accomplish this, AWACS team members must exchange, interpret and effectively weight information to optimize resource allocation decisions across team members and over time. These decisions are based on shared resources, such as strike forces (e.g. fighter/bomber aircraft), surface-to-air missile (SAM) sites, reconnaisance, refueling, and search/rescue aircraft. Relevant information must be distributed to appropriate agents and updated over time, in dynamic conditions which may require shift changes in personnel.

"..No matter how the battlefield of a particular future conflict evolves, and no matter what mix of power is used, there will always be a human in every loop, to exercise command and control.." (Ref 3). Technology has enabled rapid collection and distribution of huge amounts of information—with subsequent benefits and challenges. Information overload creates the need for mechanisms to assist decisionmakers to rapidly assimilate, interpret, and communicate relevant information to appropriate recipients. As a result there is increasing effort and focus on understanding various aspects of performance within this challenging environment—which in itself poses numerous challenges to experimental control and design.

Our approach to the study of performance within these highly complex settings begins with realistic replication of equipment used by C^3 teams and the construction of scenarios highly representative of actual task demands. In addition, subjects are operational WDs, and thus represent the training and expertise of the task domain of interest. In this study we investigated the performance of AWACS weapons directors in highly demanding scenarios, to better understand complex relationships among measures of performance, and the impact of information characteristics on decisionmaking process and outcomes. The primary research goal was to investigate relationships among measures of team effectiveness and performance and ascertain the effect of ambiguous communications on individual and team decisionmaking.

Realistic replication ensures optimal generalization of research findings to operational crew environments. However, replication alone is not sufficient to achieve either research or training goals. Scenarios must be carefully developed to manipulate constructs of interest, or to execute strategies linked to training goals identified through needs analysis. This issue is particularly salient now, as high fidelity simulations are increasingly being acquired as training technology (Ref 4). While these systems are expected to augment field exercises, it should be noted that two aspects of these systems are crucial for effective training—and are often overlooked when systems are purchased. Critical to the success of these systems, whether they be used for research or operational training are (a) systematic goal-oriented scenario development and (b) on-line and off-line performance assessment. This paper will focus on these two issues within the context of an investigative study of AWACS WD teams.

2. METHOD

2.1 *AWACS WD Simulation.*

The crewstations and scenarios which simulate the air power C^3 mission of an AWACS platform are referred to as Command, Control, and Communication Simulation Training And Research Systems (C^3STARS). The basic design of the C^3STARS research facility allows investigators to record and capture virtually every keystroke and utterance of each team member. The capability of this facility has recently been broadened by connecting the generic C^3 consoles to the Defense Simulation Internet (DSI) so that assets at other DOD facil-

ities can be integrated for multi-force simulation exercises. Scenarios are developed to investigate constructs of interest, whether they be aspects of task demand (e.g. workload, ambiguity, equipment breakdown, etc.) or aspects of individual performance (e.g. workload management under stress, assisting team members to achieve overarching team goals, strategy/tactics implementation, etc.) (Ref 5). Data captured on-line are aggregated and archived.

2.2. *Scenario Development.*

Seven scenarios were developed by our subject matter experts around a common theme. All scenarios were variations related to the central theme of conflict between two countries with a long history of hostility. Events were systematically generated to ensure equivalence in workload demands. For example, all scenarios were comprised of four 30-minute "waves" of high hostile activity. The challenge is in creating scenarios that have equivalent task demands, but are not perceived by subjects to be the same scenario. Subjects would experience several hostile aircraft at a particular time, in combination with secondary tasks, such as a search-and-rescue effort. In another scenario, at about the same time, the subjects would experience the same number of hostile aircraft, but from a different direction, with different secondary tasks. To further disguise the scenarios from recognition, names and callsigns were different in each scenario. Presentation of visual information cues must coordinate with audio cues, which can be either time- or event-driven (Ref 6).

Scenario development must also capture the team task characteristics relevant to the study (Ref 5). Our study centered upon team coordination and communication, when team members have more distributed expertise, within a context that is both demanding and ambiguous, with the same Air Defense Warning Level and rules of engagement. To create distributed functional expertise, we parceled the C^3 responsibilities of a typical air campaign of a regional conflict into functional WD roles, as typically done in the AWACS community. The three roles are (1) HVA (High value Asset) WD, (2) CAP (Combat Air Patrol) WD, and (3) STRIKE WD. The HVA WD controls the C^3 aircraft, air refueling operations, Electronic Warfare (EW) operations, and reconnaissance operations. The CAP WD controls the Defensive Counter Air (DCA) aircraft, coordinating the fire of friendly Surface-to-Air Missile (SAM) assets, and team leadership. The STRIKE WD controls the planned bombing missions, the unplanned Suppression of Enemy Air Defense (SEAD) SAMs missions, and the unplanned Theater Missile Defense (TMD) bombing missions. For a further discussion of these roles and specific examples in the context of maintaining situational awareness see Dalrymple & Schiflett (Ref 6).

Previous research (Ref 7; Ref 8; Ref 9) has demonstrated the advantages of a systematic event-driven approach to performance assessment. In our study, we embedded within each scenario 24 communication/resource allocation decision events requiring team interaction. Each event was similar in that information is presented to one teammember which should be passed to another teammember. This approach gives teammembers the opportunity to alert others to relevant information in accomplishing their area of responsibility. In each event, there is a need for (a) communication among teammembers, (b) accurate threat assessment and (c) possible resource allocation to address the threat. These events varied in the degree of ambiguity of information. Ambiguity was controlled by use of terms varying in level of certainty, from low ("estimate", "possible") to higher levels of certainty ("probable", "as stated"). Each teammember experienced 8 events where communication is initially directed to him/her, resulting in a total of 24 events per scenario. A subject matter expert observed and recorded the team actions and decisions for each embedded event.

2.3. *Data Collection.*

Data was collected on nine three-person AWACS WD teams over a six-month period in 96-97. Each team participated in seven 3-hour scenarios, scheduled over one week. The first scenario was for the purpose of familiarization with the facility and procedures. For each scenario, WDs participated in pre-mission and post-mission briefing sessions which were very similar to their usual procedures. In addition, subjects responded to a variety of instruments and measures, to collect data on personality and other self-reported characteristics/perceptions.

2.4. *Performance Measures.*

The AWACS weapons directors have two overarching goals, to (a) protect/preserve friendly assets and (b) destroy enemy assets. Both goals are served when enemy aircraft with hostile intent are destroyed. Overall outcome data are generated and interpreted in a relatively straightforward way. However, other measures of individual and team process and performance become more difficult to capture and interpret. The following diagram illustrates the tradeoffs inherent in the measurement of performance from different levels of analysis (Ref 8; Ref 10).

FIGURE 1. PERFORMANCE MEASUREMENT Hierarchy

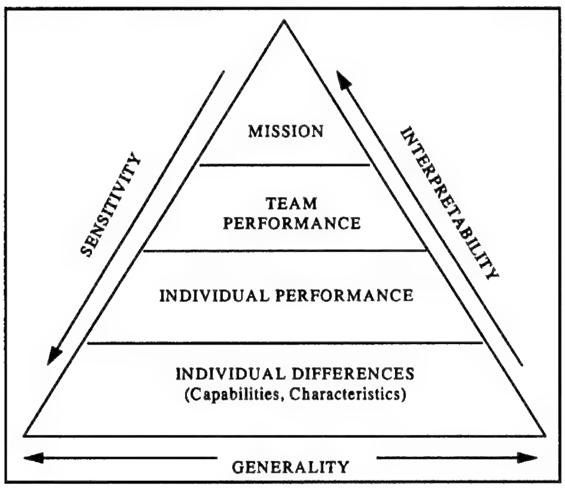


Figure 1 illustrates four broad categories of performance measurement, and some tradeoffs with regard to interpretability (performance as it relates to the mission), sensitivity (performance measures as functions of the performer as opposed to context), and generality (relationships to other teams and/or situations). At the tip of the triangle are measures related to overall mission effectiveness--those measures that indicate overall performance of the team in light of mission objectives. These relatively straightforward measures are driven by the context and goals of the team and are most easily interpreted when assessing team performance in relationship to mission outcomes.

Next are measures of team performance as defined by the degree the team as a whole managed interdependencies among team members. This level of measurement would include dimensions such as the effectiveness of communications, coordination, information exchange, and resource allocation among team members. These measures are less straightforward than mission outcomes with regard to assessing performance effectiveness. First, they may not reflect mission outcome, due to the effect of other variables, such as overwhelming number and tempo of hostile forces. The team may have coordinated information as efficiently as possible, but still fail in overall outcomes. How then, do we differentiate "good" teams from "bad" teams when all "failed" or "succeeded" in mission outcome? Keep in mind some teams may accomplish mission success in spite of poor coordination and decisionmaking. Second, there may not be a consistently positive relationship between a teamwork measure, such as information exchange, and overall performance. For example, full information exchange under conditions of high time pressure can have a detrimental effect on mission effectiveness (Ref 11; Ref 12).

Generalization of relationships among measures of team performance and mission effectiveness should be restricted to contexts having similar core characteristics of the mission context. For example, in the current investigation, measures selected to represent mission effectiveness may only be appropriate to AWACS teams, and even then, to particular types of scenarios. In this case we generated a defensive counter-air mission, where destruction of hostile aircraft is necessary to prevent loss of friendly assets. In other missions, the measure of effectiveness may be more a function of penetration into hostile territory or destruction of non-airborne hostile assets. In addition, research findings regarding effects of predictor variables or study manipulations (i.e., ambiguity, communication effectiveness) are expected to generalize to other military command and control performance domains with similar task demands, but not to team contexts which have a different profile of task characteristics.

At the bottom of the triangle are the measures of individual differences in capabilities and other characteristics. One should note that relationships among these measures and mission effectiveness will be most attenuated, due to the variety of contextual variables which can affect mission effectiveness. Instead, the impact of individual capabilities should be related to individual performance, which are then related to measures of team and mission effectiveness.

Team performance is usually expected to be related to measures of individual performance on task; however, it should be noted that the relationship is reciprocal--team performance will affect individual performance, just as individual performance affects team performance (Ref 13).

This modeling approach to performance provides greater understanding of the impact of predictors on performance. Criterion measures should be appropriate to the predictors of interest. For example, in assessing the impact of a particular drug(s) on performance, it is apparent that the greatest impact would be on measures of individual capabilities and performance, and lowest for mission effectiveness (Ref 8). The data collected in this study was also generated to be consistent with individual and team-level variables within the Multi-Level Theory of Team Decisionmaking (Ref 14). This approach is further discussed Ref 11; Ref 12). It is critical to address methodological issues with regard to conceptualization, measurement, and analysis of data when investigating relationships between individual and team-level variables (Ref 15).

In this study, we focused on ambiguity, communication and decisionmaking as they related to individual and team decision outcomes. These decision outcomes were then related to measures of mission effectiveness. Thus, most of the analyses investigated measures of team or mission effectiveness.

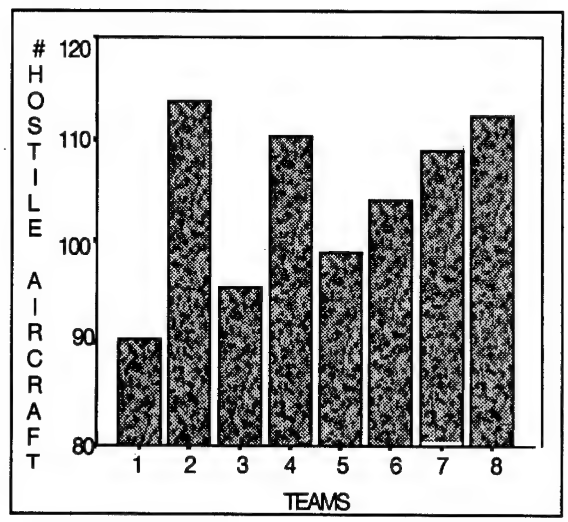
2.5. *Measurement of Mission Effectiveness.*

Mission effectiveness can be reflected by several different indices, all based on some combination or weighting of losses sustained by enemy forces in comparison to friendly losses. Certainly, failure can be clearly stated if hostile forces overtake and destroy all friendly assets and success is apparent when all hostile forces are repulsed. However, when penetration and destruction by enemy forces is not complete, mission effectiveness is usually indicated by the number of hostile aircraft/assets destroyed, in comparison to the number of friendly aircraft/assets destroyed. For example, the "kill ratio" of hostile aircraft destroyed to friendly aircraft destroyed is usually expected to be at least 3-1, to be considered effective.
Several different indices can be generated to reflect the conceptualization of mission outcome based on the number of hostile versus friendly assets which were "lost", and the reasons why they were lost (e.g. hostile action, friendly fire, fuel out..). Here, we examine three measures arising from this definition to identify the extent to which they agree—through correlational analyses and outcome with regard to ranking teams in mission effectiveness. First, we examine the most "bottomline" measure—the number of hostile assets destroyed. We then examine the number of friendly assets lost to hostile action. This number may be regarded by some as a measure of efficiency rather than effectiveness. However, it is clearly a measure of outcome as well. These two indices are then combined, in different ways, to represent overall mission effectiveness. Each measure will be described in turn, followed by analyses of interrelationship and interpretability.

2.6. *Number of hostile assets destroyed by friendly assets.*

Figure 2 describes the average number of hostile assets (in this scenario, assets are all aircraft) destroyed by friendly aircraft for each team. Despite the low number of teams, and thus low power for statistical significance, differences were significant (F = 3.82, p = 0.003). Even more impressive is the effect size (Eta = 0.64)---41% of the variance in the number of hostile aircraft destroyed is accounted for by differences between teams.

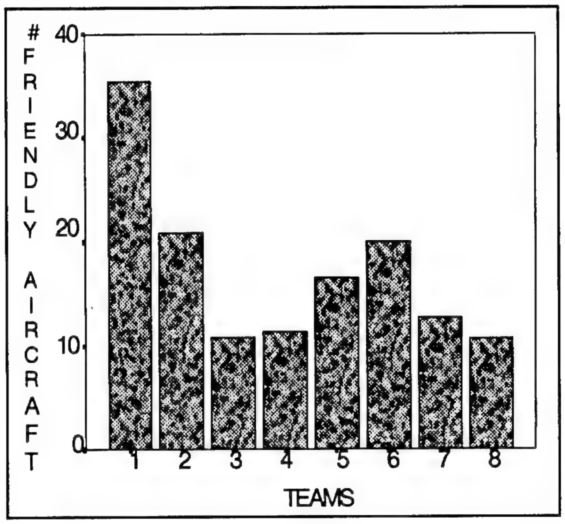
FIGURE 2. MEAN NUMBER OF HOSTILE AIRCRAFT Destroyed by Friendly Aircraft by Team.



2.7. *Number of friendly assets destroyed by hostile assets.*

Figure 3 illustrates the mean number of friendly assets for each team. Differences between teams were also statistically significant ($F = 9.21$; $p = 0.000$) and very high in effect size (Eta = .79; thus accounting for 62% of the variance in this measure).

FIGURE 3. MEAN NUMBER OF FRIENDLY AIRCRAFT Destroyed by Hostile Aircraft



While there is some correspondence between these two measures in assessing effectiveness (Pearson $r = -0.20$, $p = 0.09$), there is not complete agreement. It is clear that team 1 is most ineffective, as it had both the lowest number of hostiles destroyed, and the highest number of friendly losses. On the other hand, team 2 ranked highest for number of hostile aircraft killed (see Figure 2), but also had a relatively high number of friendly losses (Figure 3). The teams rank out differently on these two measures, particularly teams 2 and 3.

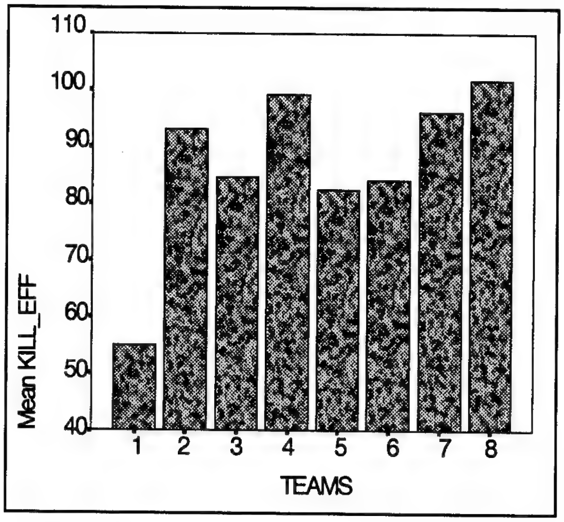
Examination of these two measures serves to illustrate that assessment of mission effectiveness can be complex and to a degree, subjective. The measure(s) chosen to represent mission effectiveness should therefore be examined in relation to alternative measures, and subjective elements should be identified and made explicit. Indeed, WD performance may be significantly affected by differing perceptions of the WD teammembers as to what comprises mission effectiveness. For example, to what degree should destruction of hostile forces be weighted against loss of friendly assets, including personnel? As another example, should an error of the "false alarm" variety (i.e. destroying a nonthreatening "target") be punished to a greater degree than an error in the opposite direction (i.e. not destroying a threatening target).

2.8. *Kill Efficiency and Kill Ratios.*

If we turn to military doctrine for interpretation of mission effectiveness, we find that indices of effectiveness will differ according to the overall scenario, rules of engagement, and threat level. As a general rule of thumb, the ratio of hostiles killed compared to friendlies killed is usually expected to be around 3 to 1. Our domain experts assure us that the number of hostiles killed is the more important number, by about a factor of three. To avoid problems with the use of ratio numbers (possible division by 0) we generated a score, referred to as kill efficiency, that is a function of the number of hostile assets destroyed by friendly assets minus the number of friendly assets destroyed by hostile assets. However, if we generate this number comprised of the number of hostile assets lost and weight this number by three, minus the (unweighted) number of friendly losses, we find that the resulting number correlates very highly with the number of hostiles killed, such that the consideration of friendly losses really does not make a difference in the measure. This is particularly salient when ascertaining relationships with other variables. Thus, to generate an indice that is actually a function of both measures, we chose to have our measure of kill efficiency be based on equal weighting of both, that is, the number of hostile losses (due to friendly action) minus the number of friendly losses (due to hostile action).

Figure 4 illustrates team performance when assessed by this measure of kill efficiency. Again, a different ranking of teams results. In essence, this measure averages the difference in rank generated by the two measures used previously.

FIGURE 4. MEAN KILL EFFICIENCY BY TEAM



2.9. *Kill Ratio.*

If we divide the number of hostile aircraft destroyed by friendly aircraft by the number of friendly aircraft destroyed by hostile aircraft, we attain yet another representation of mission outcome, which was feasible in this case because none of the scores involved division by 0. This measure is independent of the specific number of aircraft lost, per se. Thus, if one team killed 20 hostiles and lost 2 friendlies, they would have the same kill ratio as one which killed 30 hostiles and lost 3 friendlies. In contrast, using the kill efficiency score, the team that killed 30 hostiles would be assessed as more effective. Again, teams will rank out differently using this measure, compared to the kill efficiency measure. Teams such as team 2, which killed many hostile aircraft but also lost a higher number of friendlies, ranked higher using the kill efficiency score as compared to the kill ratio.

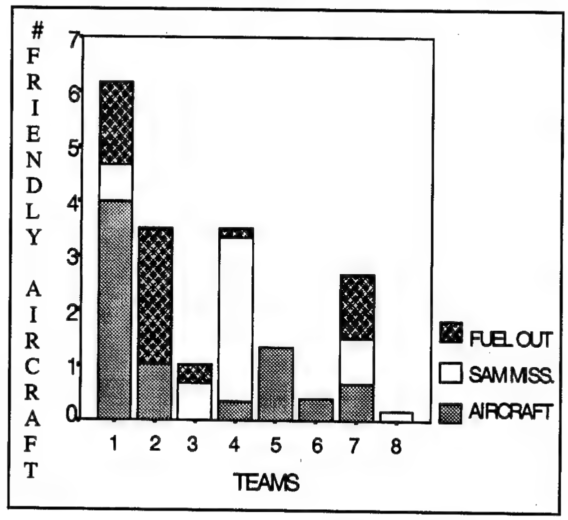
2.10. *Outcomes weighted by "points".*

Subject matter experts assigned points to represent the relative value of different types of assets, such as Airbases (50 points) and aircraft (F-15 = 15 pts; E-3 = 50 pts; Mig-25R= 40 pts, etc.). Using this weighted approach, we calculated an outcome score based on the weighted value of hostile assets lost (for any reason) minus the weighted value of friendly assets lost. The pattern of team rankings and correlations with other variables reveal a pattern similar to non-weighted scores. This is consistent with the robust nature of linear equations comprised of several variables—variables can vary a great deal in relative weights yet result in very similar functions with other variables (Ref 16).

2.11. *Friendly aircraft lost due to non-hostile action.*

Variables in this category include the number of friendly aircraft lost due to fuel out, shot down by a friendly SAM missile, and the number of friendly aircraft shot down by friendly aircraft. Figure 5 describes these measures for each team. It can be seen that team 1 performed very poorly on this measure. Other teams that had higher losses due to friendly action differed on team rank based on other indices. For example, team 2 incurred a relatively high loss of friendly aircraft due to friendly fire or fuel out. However, it had the highest number of hostile aircraft destroyed, and a relatively high kill efficiency score. On the other hand, team 6 had a low incidence of friendly fire, but a higher number of friendly aircraft destroyed by hostile forces, and a relatively less effective kill efficiency score.

FIGURE 5. FRIENDLY AIRCRAFT LOSS DUE TO Friendly Action: By Type of Action and By Team



3. MEASURES OF TEAM FUNCTIONING

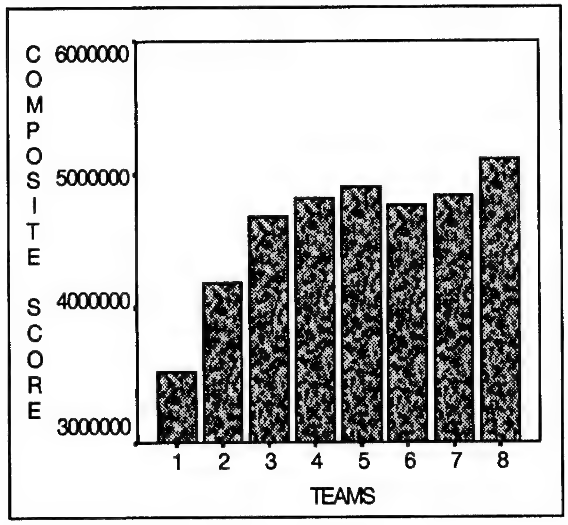
3.1. *Composite Score.*

The previous section demonstrates that even simple quantifications of outcome can be problematic. Contradictory assessments can arise from different measures and there is an element of subjectivity in both choice and weighting of outcome measures. Subject Matter Expert (SME) judgments, on the other hand, may (or may not) be generated with a high degree of subjective certainty and inter-rater agreement. It is reasonable to expect that SME judgments are based on more factors than a simple count of losses. However, capturing and quantifying the underlying "policy" of these judgments is difficult. The number of possible factors is high, and the number of decision events is low, making regression analysis inappropriate even if all factors could be identified. To explore this issue, we generated an a priori indice based on WD expertise and judgment with regard to relative importance of indice components. The indice was generated by a former WD instructor and mathematics major, and is a summation of the following variables:

- (a) # friendly assets not destroyed
- (b) # friendly assets destroyed
- (c) # hostile assets destroyed by friendly action
- (d) Kill Ratio (Sin(arctan[HACd/FACd]) x HACd)
- (e) Air refuelings completed
- (f) Assign/defer actions completed (transfer responsibility for resource)
- (g) # friendly aircraft lost to fuel out
- (h) # friendly aircraft lost to friendly SAM missiles
- (i) # friendly aircraft lost to friendly aircraft fire
- (j) # hostile aircraft jammed
- (k) # friendly aircraft jammed
- (l) friendly penetration of hostile area (maximum distance into hostile territory)
- (m) hostile penetration of friendly area
- (n) hostile aircraft lost to fuel out.

Figure 6 provides mean composite scores for each team. Again, it can be seen that teams will rank out somewhat differently, especially teams that are average. Team 1 is consistently poor regardless of outcome measures, and that is demonstrated with this composite score as well. There is somewhat less obvious variance among teams on this measure, and in general, this measure appears to coincide with the mean number of friendly aircraft destroyed by hostile—except in reverse ranking. Teams that had a high number of friendly aircraft destroyed have a low number in terms of composite score, and vice versa.

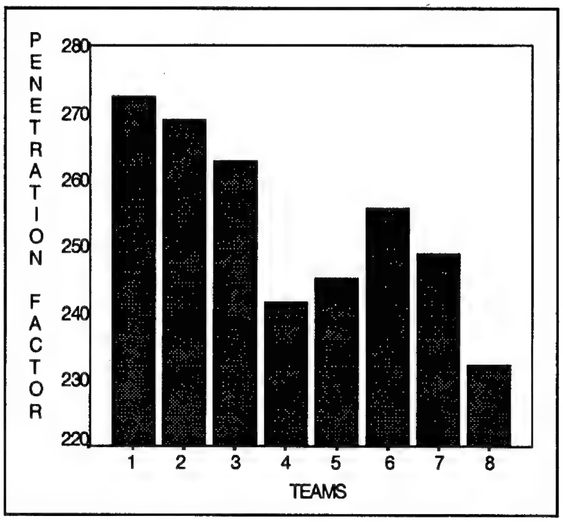
FIGURE 6. MEAN COMPOSITE SCORE BY TEAM



3.2. *Penetration into Territory.*

Another indice of team function, as identified by SME input, is the extent to which hostile forces penetrated friendly territory, and vice versa. These scores were generated, and correlated with other indices of mission and team performance. There appeared to be little variance between teams on penetration of hostile territory; however, penetration of friendly territory by hostile forces did differ significantly among teams ($F = 2.35$; $p = 0.05$), and is portrayed in Figure 7. It can be seen that team 1, which scored poorly on other indices of mission outcome, also scored poorly on this measure, whereas team 8, which had the highest mean kill efficiency scored, also scored relatively well on this measures. However, team 2, which also had a high kill efficiency score, did not score well on this measure.

FIGURE 7. MEAN PENETRATION OF FRIENDLY Airspace by Hostile Forces, By Team



3.3. *Team Communication.*

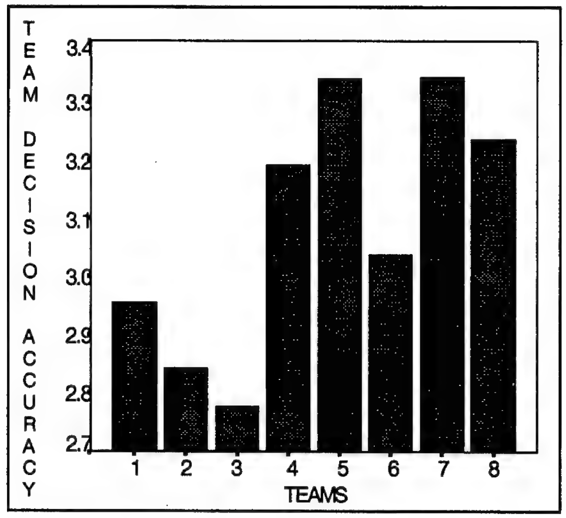
Team communication was measured through observational data regarding embedded communication/decisionmaking events. Teams participated in six two-hour sessions, each of which had 24 decision events, for a total of 144 decision events for each team. For each decision event, information was provided by voice to one team member, which was relevant to another team member. An SME observed and recorded whether the initial message was acknowledged by the crewmember, and then transferred to the appropriate team member; whether the teammember receiving that message acknowledged it and what that individual decided to do as a result; what the team decided to do, and any special circumstances.

A drawback in testing the prediction that information transfer (from receiver to relevant member) would affect decision accuracy was the lack of variance in the predictor. Teammembers nearly always (83% of the time) passed the information to the appropriate person. While reassuring to know team members are competent under these circumstances, the lack of variance is likely the reason for lack of relationship between information transfer and decision accuracy.

3.4. *Team Decision Accuracy.*

Individual and team decision accuracy was assessed through comparison of a priori SME judgments of optimal action. For each decision event, a SME provided judgments of the best course of action. These judgments were provided by two SMEs, agreement was high, and any differences were resolved.

FIGURE 8. MEAN TEAM DECISION ACCURACY BY Team



Measure of accuracy were based on the difference between actual and a priori judgments, such that a match with the optimal a priori judgment was assigned a point value of "4", a match with the 2nd best a priori judgment was assigned a "3", a match with the 3rd best a priori judgment was assigned a "2", a match with the 4th best a priori judgment was assigned a "1", and no match was assigned a "0". Thus, the higher the score, the more accurate the judgment. The following section describes relationships among these measures of team outcomes and team effectiveness.

3.5. *Intercorrelations.*

Tables 1, 2, and 3 provides descriptive statistics and intercorrelations for all variables described above. Examination of these relationships illustrates the need to examine an array of outcome measures before attempting to choose a single representative measure of mission effectiveness. In addition, any composite scores should be developed based on rational means; however, functional relationships to component scores should also be examined.

Examination of correlations reveal several patterns. We focused on the relationships of three indices of team effectiveness, kill efficiency, the composite score, and team decision accuracy, to measures of mission effectiveness (see Table 3). Note the differences among these measures in their pattern of relationships. As expected, Kill efficiency was highly related to both hostile and friendly losses, since it is comprised of these two variables. On the other hand, the composite score was unrelated to hostile loss, but highly related to friendly loss by hostile aircraft and by friendly fire. In contrast, the mean team decision accuracy was unrelated to the number of friendly aircraft destroyed but significantly related to the number of hostile aircraft destroyed (by friendly aircraft). The decision events themselves were based on threat assessment and resource allocation to threat, thus, effective decisionmaking in these events should result in effective destruction of hostile assets. It was surprising to find that this measure of team decision accuracy was not related to another measure, the overall composite score.

It is clear that the composite score is primarily a function of the number of friendly aircraft destroyed. This was expected by those who generated the algorithm. The rationale is based on effectiveness over time. If many hostile assets are destroyed at high cost to friendly assets, one loses the capability to continue over time. In addition, friendly losses can quickly result in loss of political support, such that the conflict is unsuccessful in spite of mission accomplishment (in terms of "bombs on target").

TABLE 1. DESCRIPTIVE STATISTICS

	Mean	Std. Deviation	N
# Hostile Assets destroyed by Friendly Assets(HACd)	104.23	12.88	47
# Friendly Assets destroyed by Hostile Assets(FACd)	17.21	10.09	47
Kill Efficiency/Effectiveness(KE)	87.02	17.89	47
Kill Ratio(KR)	8.10	5.03	46
#Friendly Aircraft destroyed by Friendly Aircraft(F_FR)	0.98	2.16	47
#Friendly Aircraft lost to fuel out(F_FU)	0.72	1.53	47
#Friendly Aircraft destroyed by Friendly Missiles(F_FSS)	0.68	1.25	47
Composite Score(CS)	4.59E+06	7.36E+05	47
Friendly Penetration of Hostile Territory (FPF_ALL)	790.89	50.02	47
Hostile Penetration of Friendly Territory (HPF_ALL)	253.36	24.92	47
Mean Team Decision Accuracy (TDA)	3.02	0.42	48

TABLE 2. INTERCORRELATIONS

	HACd	FACd	KE	KR	F_FR	F_FU	F_FSS	CS	FPF	HPF	TDA
HACd	1.000	-.20*	.83***	.44***	-.19*	-.03	.25**	.07	.20*	-.14	.50***
FACd		1.000	-.71***	-.76***	.43***	.34**	-.08	-.66***	.31**	.44***	-.15
KILL_Efficiency			1.000	.70***	-.38***	-.21*	.22*	.42***	-.03	-.34***	.44***
KILL ratio				1.000	-.31**	-.31**	.16	.44***	-.08	-.44***	.45***
NF_FR (friendly fire)					1.000	.37***	.102	-.51***	.27**	.16	.05
NF_FU (friendly fuel out)						1.000	-.08	-.45***	.11	.20*	-.05
NF_FSS (friendlies lost by friendly SAM sites)							1.000	-.08	.09	.14	.10
Composite Score								1.000	-.16	-.36**	.08
FPF (friendly penetration of hostile territory)									1.000	.068	.23*
HPF (hostile penetration of friendly territory)										1.000	-.44***
TDA (team decision accuracy)											1.000

significant at p<0.10; ** significant at p<0.05, *** significant at p<0.01 (one tailed tests).
N= 47 for all variables except TDA (N=27)

TABLE 3. COMPARISON OF 3 INDICES OF TEAM EFFECTIVENESS

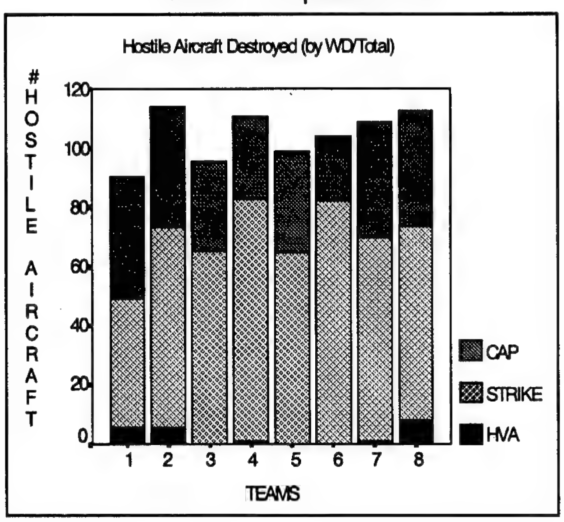
	FACdes	HACdes	FAC_fuel	FAC_fr	FPF	HPF
Kill Efficiency	-.71***	.83***	-.21*	-.38***	-.03	-.34***
Composite Score	-.66***	.06	-.45***	-.50***	-.16	-.36***
Team Dec Accuracy	-.15	.50***	-.05	.05	.23*	-.44***

3.6. *Team performance is not necessarily individual performance.*

Measures of team performance should not be generalized to represent the performance of individual team member, since the relevance of team measures to individual performance is affected by the roles, responsibilities, and interdependencies of the team. Team success or failure may be a function of the performance of a single member. In fact, team performance outcomes may not be related to the performance of any of the team members, but may instead be a function of circumstance. For example, a team may succeed or fail in spite of the performance of team members.

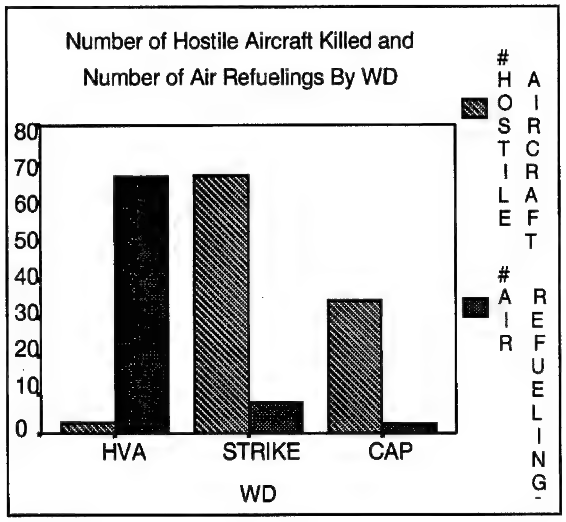
To illustrate this point, Figure 9 depicts the number of hostile aircraft destroyed by each WD position within each team, averaged across the six scenarios. Results indicate differences by WD. WD1 had very few hostile aircraft kills, while WD2 consistently had high numbers.

FIGURE 9. NUMBER OF HOSTILE AIRCRAFT destroyed by Team and by Team member role specialization



The next graph more clearly illustrates differences due to WD role. The WD2, in charge of strike functions, had higher responsibility for attack of hostile aircraft, and thus higher numbers are associated with WD2. In contrast, WD1 was in charge of high value assets, and thus more responsible for air refuelings.

FIGURE 10. NUMBER OF HOSTILE AIRCRAFT destroyed by Friendly Aircraft and Number of Air Refuelings by Teammember Role.



4. DISCUSSION

This report described the indices of performance used in our study of AWACS performance in a realistic war simulation. The development of valid and reliable measures of team performance is critical to assessment of operational performance and to the development of selection and training systems to improve team performance. Our approach begins with the identification of performance measures which reflect team outcomes (mission effectiveness). In addition, measures of team process are generated from theory, past research findings, SME experience, and empirical evidence. Measures of team performance are distinguished from measures of individual performance.

In this study, we focused on the measures of mission and team effectiveness. These measures are fundamental to effective performance modeling, and can be problematic in complex dynamic environments. While quantitative measures of team outcomes can often be easily attained, we show that subjective elements must be identified and resolved. In this study we described several measures of mission effectiveness, we demonstrated how these measures can result in different assessments of team effectiveness, and we discussed assumptions that underlie these measures. These assumptions should be made explicit, and measures chosen that are consistent

with actual military doctrine. At the same time, measures should be chosen that are most consistent with predictions inherent in the study. In most cases, an array of measures should be captured as different aspects of performance, thus allowing greater clarification and understanding of the impact of interventions on mission performance.

Our approach also relies on systematic goal-oriented scenario development to ensure elicitation of the behavioral constructs of interest. In this study, team communication and decisionmaking were of primary interest, therefore decision events were systematically generated. The dynamic nature of the C^3 task challenges the creation of independent events, making a priori specification of correct action difficult. In other words, teams may be forced to take action other than what was pre-specified as optimal, wherein that action may be the optimal action, given circumstances at that time. A follow-up analysis of this data is described in a related paper in this conference (Ref 17).

The choice of measure to represent mission outcome should be driven by scenario characteristics, such as type of strategy (offensive, defensive), type of tactics, prevailing doctrine, available assets, and tactics/assets used by hostile forces. Investigators must work closely with subject matter experts to select outcome measures that are appropriate to the scenario dynamics and consider the relative importance of outcomes such as asset loss (relative importance of assets), penetration of territory, and/or fratricide.

5. REFERENCES

1. Ripley, T., 1998. http://www.global-defence-review.com/AirbourneWatch.html

2. Boyne, W. J., "Gulf War - A Comprehensive Guide to People, Places & Weapons." Signet., 1991.

3. Air Force Scientific Advisory Board's New World Vistas., 1997. http://web.fie.com/htdoc/fed/afr/sab/any/text/any/vistas.htm

4. Clark, C., "Simulation Buying Spree Needs Restraint", Defense Week, June 9, v18., June 1997.

5. Dalrymple, M. A., "Evaluating Airborne Warning and Control System Strategy and Tactics as they relate to Simulated Mission Events.", Technical Paper (AL-TP-1991-0049), 1991.

6. Dalrymple, M. A., & Schiflett, S. G., "Measuring Situational Awareness of AWACS Weapons Directors.", Paper presented at the Symposium on Situational Awareness in the Tactical Air Environment, Patuxent River, MD, June 1996.

7. Eddy, D., "Selected team performance measures in a C3 environment: An annotated bibliography.", USAFSAM (USAFSAM-TR-87-25), 1989.

8. Dalrymple, M., Eddy, D., & Schiflett, S., "Comparative Effects of Antihistamines on AWACS Weapons Director Performance under Sustained Operations.", Final Technical Report., (F41624-91-C-2008), 1995.

9. Fowlkes, J., Lane, N., Salas, E., Franz, T., & Oser, R., "Improving the measurement of team performance: The TARGETs methodology.", Military Psychology, 6, pp. 47-61, 1994.

10. Schiflett, S., Strome, D., Eddy, D., & Dalrymple, M., "Aircrew Evaluation Sustained Operations Performance (AESOP): A triservice facility for technology transition.", Technical Report (USAFSAM-TP-90-26), 1990.

11. Elliott, L. R., Neville, K., Dalrymple, M., & Tower, S., "Impact of Ambiguity on AWACS Team Communications and Decisionmaking.", Paper presented at the Ninth International Symposium on Aviation Psychology, Columbus, OH., 1997.

12. Elliott, L. R., Dalrymple, M., & Neville, K., "Assessing Performance of AWACS Command and Control Teams.", Proceedings from the 1997 Meeting of the Human Factors Society, September, 1997.

13. Regian, J. W. & Elliott, L. R. (In Review). "Intelligent Training for Team Performance: Theory-based Development of an Enhanced Platform for Team Training and Performance Research.", (Technical Report in Review).

14. Hollenbeck, J. R., Ilgen, D. R., Sego, D., Hedlund, J., Major, D., & Phillips, J., "Multi-level theory of team decision making: Decision making in teams incorporating distributed expertise." Journal of Applied Psychology, 80, 2 pp. 292-317, April, 1995.

15. James, L. R., "Aggregation bias in estimates of perceptual agreement." Journal of Applied Psychology, 67, 2 pp. 219-229, 1982.

16. Dawes, R. M., "The Robust Beauty of Improper Linear Models in Decisionmaking.", American Psychologist, 34, pp. 571-582, 1979.

17. Cobb, G., Mathieu, J., Elliott, L. R., & Dalrymple, M., "The Impact of Environment Complexity and Ambiguity on AWACS Crew Performance in Simulated Combat Scenarios.", Paper to be presented at the 1998 NATO Symposium on Collaborative Crew Performance in Complex Operational Systems, Edinburgh, United Kingdom, 20-24 April, 1998.

Group Operations: Synergy Through the Exploitation of Technology and Tactics

A.P. Hall, Operational Analysis Dept (W392A)
British Aerospace Military Aircraft & Aerostructures
Warton Aerodrome, Warton, Preston PR4 1AX, UK

T. H. Normanton (Smiths Industries Aerospace, Cheltenham, UK)
B. G. Smith (GEC-Marconi Electro-optics, Basildon, UK)

SUMMARY

This paper examines Group Operations (i.e. missions and composite air operations containing more than one aircraft of similar or different types). In particular, it considers current practice, looks at key issues for consideration, and identifies future potential from operational, technical, and cost effective viewpoints.

From these operational requirements, functional requirements are derived and grouped into three broad areas: improved situation awareness; improved reactive capability; and improved co-ordination.

Technologies and 'Total System' implementation concepts to meet such requirements are then identified. The paper concludes with a summary of the key technologies, their benefits and associated risks. Furthermore, the possibility is proposed for an evolutionary route to achieving an enhanced Group Operations capability with near, medium, and long-term options. However this can only be achieved through detailed Technology Feasibility studies and Tactics development.

1. INTRODUCTION

In today's post Cold War era, the RAF is faced with a number of capable, smaller, more diverse threats. In recognition of this, NATO's New Strategic Concept [Ref 1] calls for a capability to counter a major threat and intervene in smaller crises or undertake peace support operations around or outside NATO's borders. These facts, when combined with shrinking defence budgets, pose a formidable challenge for the RAF. In short, it must 'do more with less'. A point echoed by Gp Capt. Woolley from the Air Warfare Centre [Ref.2]:

> *"The RAF's former role in NATO's Central Region has left us with many strengths, including a qualitative operational 'edge' over most potential adversaries. However, now that we can more clearly see the emerging trends of future operations, perhaps the time has come for a reappraisal of how we should organise, equip, and train".*

Of course the changing world scene in political, industrial and technological terms presents similar challenges for the defence industry. It is the way that the RAF and the defence industry meet these challenges that will determine their future success. In particular for group or COMposite Air Operations (COMAO), this can be achieved through the exploitation of new and emerging technologies and appropriate tactics. For example, technologies such as covert communications and distributed data fusion can enable increased synergy and co-operation between a group of aircraft.

This paper examines existing operational implementations and future concepts for both air-to-air and air-to-ground group operations. Furthermore, it identifies and recommends future additions and/or modifications to avionics systems to realise such concepts.

1.1 Terminology

The term 'Group Operations' per se is ambiguous. To some it is seen as close co-ordination between aircraft in a small formation (e.g. a four-ship); to others it represents co-ordination (usually pre-planned, pre-mission) across a large raid package. In short, individuals have different perceptions of Group Operations.

The term 'Group' may be applied at different levels to the following:

- A collection of like aircraft with the same objective;
- A collection of like aircraft with a composite objective;
- A collection of like aircraft with physically or temporally disparate objectives;
- A collection of unlike aircraft in the same role;
- A collection of unlike aircraft co-operating in unlike roles;
- A joint and/or combined force involving air/land/sea assets.

The RAF however uses more precise terminology to describe 'Group Operations', as the following (paraphrased) examples illustrate:

Sortie: A single aircraft performing a task in pursuit of an objective.

Mission: More than one aircraft (usually of the same type) pursuing the same objective.

Composite Air Operations: More than one aircraft (of similar or different types) pursuing the same overall objective, possibly with different individual sub-objectives. For example, a bomber raid escorted by fighters. Alternatively, a mixed package of fighters (e.g. Tornado F3s with F-15Cs) on a Fighter Sweep mission.

We adopted this terminology for reasons of clarity and consistency. Furthermore, to maintain sight of the aims and objectives, a suitable working definition or vision of a future 'group operations' capability has been derived:

> *"An enhanced capability to perform missions and composite air operations through the exploitation of current and future technology with appropriate use of tactics."*

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

2. CONTEXT

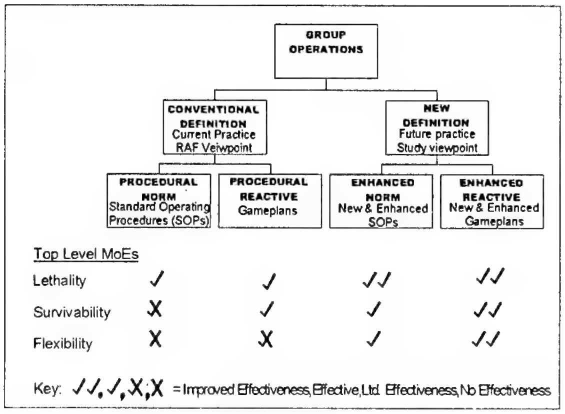


Figure 1: Group Operations in Context

Figure 1 places the RAF and our viewpoints of 'Group Operations' in context. The conventional definition relates to current RAF practice. Here, current/in-Service technology is used with operational and tactical doctrine, techniques, procedures (both normal & emergency) and training to deliver a 'Group Operations' capability. The new definition relates to possible future RAF practice. Where both current and future technology are used in conjunction with new and appropriate doctrine, procedures and training to deliver an 'Enhanced Group Operations' capability.

Note, the measures of effectiveness (MoEs) shown in the diagram broadly illustrate where potential improvements could be made and to what extent.

Current RAF practice makes best use of available technology and resources and can be characterised as:

Procedural/Prescriptive

The conduct of missions and composite air operations is largely governed by techniques and procedures in the form of Operational Orders and Standard Operating Procedures (SOPs). These exist at various levels - Squadron, Wing, or Force (e.g. Harrier Force) - in order to provide (at the appropriate level):

- Standardisation (to reduce ambiguity);
- Inter-operability (to improve availability/effectiveness);
- Conciseness (to reduce complexity).

Limited Communication

In medium and high intensity conflicts, communication between both airborne assets and with C^3I assets is strictly limited by emission control procedures for reasons of surprise and security. Where radio silence is broken, as in the case of an emergency or unexpected engagement, voice messages are normally conveyed in short bursts using standard codewords.

In low intensity conflicts and operations other than war such restrictions may be relaxed, but this is largely scenario dependent. Furthermore, there is a strong emphasis on Rules of Engagement in such operations and tight political control is often the norm. Consequently, such operations are typically characterised by significant amounts of communication, cross-checking, and co-ordination; a situation that is often exacerbated when placed in a multi-national context.

Limited pro-active/reactive capability

Responses to changed circumstances, particularly unexpected engagements by enemy fighters or SAMs, is largely reliant upon pre-briefed gameplans. These can be generic or specific to a 'known' threat type and provide a coherent and standard set of reactions. However, they are by no means comprehensive. This practice, when combined with limited communication, effectively limits the RAF's ability to exploit some of the key 'Principles of War' [Ref 3] during missions and composite air operations, namely:

- Rapid concentration (and dispersion) of Force at the decisive time and place;
- Economy of effort (i.e. balanced/optimum use of resources);
- Flexibility;
- Co-operation.

Prior to and during the Gulf War (c.1991), air power in general may have operated within a somewhat rigid structure. For example, if a raid package required support assets (e.g. SEAD aircraft) both elements would be time co-ordinated but otherwise independent. It is now perceived that there is scope for improving such a situation. Through increased co-ordination between airborne assets and associated C^3I.

2.1 Current Composite Air Operations

Based on discussions with the Tactical Doctrine and Training element of the RAF's Air Warfare Centre, this section describes a typical current day COMAO as shown diagramatically in figure 2, appendix B.

The target is assumed to be a large airfield located approximately 100 km beyond the Forward Line of Own Troops (FLOT). The COMAO package is composed of the following elements:

Fighter Sweep

A number of fighters flying ahead of the main raid package, typically at medium altitude, with their radars active and tasked with clearing a pre-defined corridor of airborne threats.

Main raid package

A main raid package of variable composition which is dependent upon the mission objective and perceived threat.

It may contain the following:

- Embedded fighters (i.e. within or to the side of the package performing an escort function);
- Embedded Suppression of Enemy Air Defences (SEAD) aircraft (usually at the front of the package);
- Interdiction aircraft.

Note, the mission commander will usually be part of the main raid package and may delegate 'kick authority' (i.e. re-routing control) to other elements within the COMAO e.g. SEAD aircraft.

Supporting Elements

Several supporting elements operating outside the main raid package typically engaged in:

- Reconnaissance:

  Pre-raid: Up-to-date targeting information is passed to the main raid package via a Reconnaissance Attack Interface (RAI).

  Post-raid: Facilitates a Battle Damage Assessment.
- SEAD;
- Tankers;
- AWACS.

Flexibility of current Composite Air Operations

Since the size of a COMAO may be anything up to 50 aircraft there is only limited flexibility to respond to changed or unforeseen circumstances. Where appropriate, minor changes to the route or Time-On-Target (TOT) can be made. However, these pose significant situation awareness and co-ordination problems amongst the different elements of the COMAO. For example, if the main raid package calls a late TOT, the preceding SEAD and Fighter Sweep must be notified to ensure adequate protection is available over the target area at the now later time. If not the mission must be aborted. Ultimately, it is the mission commander who must exercise judgement, balancing overall success against the risk of conflicting with other concurrent missions in the area.

2.2 Future World Trends

At a time of significant political, industrial and technological change it is perhaps opportune to consider how current RAF practice may evolve. This section identifies several key issues that must be addressed if the RAF is to meet the requirements of a changing world scene.

Out of Area Operations

Increasingly the RAF is being involved in Out of Area (OOA) operations, often to undertake Peace Support Operations (PSO). These are typically Joint (Multi-Service) and Combined (Multi-National) operations and may be undertaken by NATO, WEU, or UN forces. Such crisis or conflict resolution activity may involve a variety of operations ranging from disaster relief, peacekeeping and peacemaking, to peace enforcement.

Problems associated with multi-national peace support operations are not to be underestimated. There are a number of difficult areas, particularly when non-NATO nations are involved, namely: Command and Control (C^2); Rules of Engagement; Communications inter-operability.

Changing nature of the threat

As Gp Capt. Woolley [Ref 2] points out:

> *"Peace Support Operations (PSOs) will increasingly influence how we (the RAF) operate, organise, equip and train."*

Nevertheless, the changing nature of the more sophisticated threats means that War operations will also be affected. Such threats now have advanced technologies such as stealth, low frequency radars, and directed energy weapons, all of which pose a formidable problem.

New & emerging technologies

To counter such threats, it will be necessary to exploit new and emerging technologies and develop new and appropriate tactics. For example, warning and responses to threats can be improved through sensor management across a group of aircraft, covert communication, data fusion, and advanced cockpit-vehicle interfaces (CVI). US projects such as TALON SWORD and TALON LANCE which aim to achieve enhanced situational awareness by augmenting aircraft sensors with information from space-based assets and other aircraft.

Future combat aircraft

Future combat aircraft must be designed to meet the exacting requirements of both peace support and wartime operations. Furthermore, such designs must exploit both current and emerging technologies to ensure efficient and effective missions and COMAO. Inter-operability is likely to be a major concern for such an aircraft, on two counts:

- Multi-national operations may require 'backwards' compatible systems to enable future combat aircraft to operate with less sophisticated , non-NATO standard aircraft;
- Different aircraft build-standards (across a force of aircraft of the same type) may lead to sub-optimum effectiveness.

2.3 Mission Analysis

Whilst the key issues described above are important, their broad nature precludes a detailed investigation of future COMAO. Consequently, we chose here to examine future COMAO in the context of the Future Offensive Air System (FOAS). This weapon system is likely to be multi-role: primarily air-to-surface with a credible, albeit secondary, air-to-air role. The final solution has yet to be defined but is likely to be an optimum mix of manned and unmanned combat air vehicles and stand-off weapons. As such FOAS serves as a useful baseline from which to examine COMAO.

Key design design drivers for FOAS are:

- Affordability;
- Lethality;
- Flexibility;
- Availability;
- Survivability.

Within this context, we considered:

- Strategic Operations;
- Offensive Counter Air (OCA) Operations:

  SEAD;
  Fighter Sweep (FS);
  Escort;
  Airfield Attack.
- Air Support of Land Operations:

  Air Interdiction (AI);
  Offensive Air Support (Battlefield Air Interdiction (BAI));
  Tactical Air Reconnaissance (TAR).
- Maritime Air Operations:

  Anti-surface Warfare (ASUW).

Operational analysis of these missions was focussed on two critical phases within each mission: Target Attack and Threat Penetration(i.e. both Surface-to-Air and Air-toAir). Each mission was analysed and key operational issues relating to COMAO identified. For example:

2.3.1 Target Attack

Air Interdiction (Man-in-the-loop)

Target type:	Fixed hard targets Single point e.g. Bridge.
Weapon type:	Man-in-the-Loop Guided Bomb e.g. Paveway III, AGM-130.
Key issues:	Highly organized weapon aiming: coordinated acquisition & targeting; Possible third party targeting.

Fighter Sweep

Target type:	Highly mobile & threatening e.g. enemy fighters.
Weapon type:	Autonomous Medium & Short Range Missiles, Gun, DEW.
Key issues:	Coordinated target acquisition; Weapon-to-target allocation; Need for high situation awareness & mutual support; Need to maintain cohesion.

2.3.2 Threat Penetration

Threat Avoidance

Threat type:	Enemy SAMs and/or Fighters
Key issues:	Acquisition of threats; Mission profile (vs. SAM); Group tactics: highly defensive, no weapons just countermeasures; Need for high situation awareness & mutual support; Need to maintain cohesion.

3. FUNCTIONAL REQUIREMENTS FOR GROUP OPERATIONS

For future COMAO, technology alone will not deliver an enhanced capability. Sound and coherent tactical doctrine must be in place. Furthermore, for full exploitation, appropriate training policies and sound teamwork principles must be followed to ensure aircrew competence and confidence in the use of new technology. So enhanced in-service capability can only be achieved through a combination of: technology; appropriate tactics and effective training.

From our analysis (which employed teamwork [Ref 5] and systems theory [Ref 8]), the following key improvements in functionality were identified:

- Situation awareness;
- Reactive capability;
- Co-ordination.

3.1 Improved Situation Awareness

Situation awareness is a term used to describe the view perceived by a person or object of itself, the outside world or external environment, and the relationship between them. In essence, it is the first two phases of Boyd's Observation, Orientation, Decision and Action (OODA) cycle, details of which are provided in appendix A. Nevertheless, we considered that the term 'Situation Awareness' could be decomposed into three specific categories:

- High Situation Awareness;
- Common Situation Awareness;
- Role Situation Awareness.

High Situation Awareness

High situation awareness relates to the conventional views of maximum sensor coverage, high accuracy, and fast update rates. The aim here is to shorten and improve the quality of the COMAO's Intelligence gathering (i.e. Observation and Orientation) phase. This facilitates quick and effective action and helps to generate a superior tempo of operation relative to a threat or target.

With the advent of JTIDS, this form of situation awareness will undoubtedly improve. However, even JTIDS has its limitations: in particular, data latency (typically 12 seconds) and unnecessary multiple reports (from different observers) of the same contact. This leads to stale and multiple contacts, all of which must be resolved to enable an accurate situation assessment.

For future COMAO, therefore, the following functional improvements to high situation awareness are required:

- Optimisation of the COMAO's sensor coverage, update rate, and accuracy;

Sector allocation to individual COMAO package members, for example, could enable the effective 'dwell time' in any one sector to be increased, resulting in improved detection and tracking performance.

Alternatively, all but one of the COMAO package members could use passive sensors for threat/target detection. The 'active' member would use low probability of intercept (LPI) radar techniques to covertly gather accurate threat/target range data which would then be disseminated to the rest of the COMAO.

- Time tagging of sensor reports.

  By time tagging sensor reports (i.e. target/threat detections) as and when they occur, and distributing such information throughout the COMAO, problems associated with data latency and multiple contacts can in many instances be resolved.

Of course, whilst all members of a COMAO may possess high situation awareness, they may not necessarily have a common 'big picture'. This is the second category of situation awareness.

Common Situation Awareness

Common situation awareness relates to the view that 'man is the limiting factor' to situation awareness. In short, technology provides the data, and then transforms and presents it as information. Each member of the COMAO will then draw conclusions from this based upon their individual perception of the situation. Such perceptions are often incongruent and biased towards a member's own objectives and circumstances. In many instances, this can lead to reduce synergy and cohesion amongst the COMAO.
There is a requirement, therefore, for each member of the COMAO to have common data about the organisation and objectives of the COMAO, and the environment in which the package is operating. This data will be transformed and presented as information, so that each member may form a common perception of the total group situation, as well as the situation applicable to his own objectives and circumstances.

In summary, common situation awareness is gained through common perception; see figure 3.

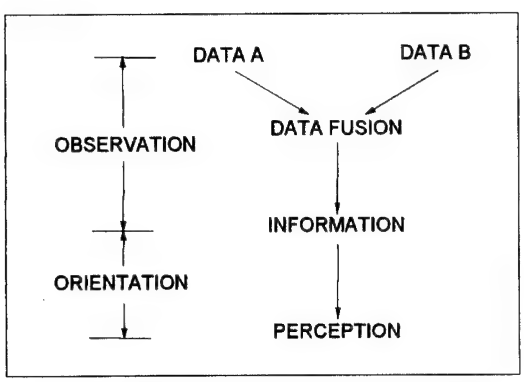


Figure 3: The Elements of Common Situation Awareness

Role Situation Awareness

This third category is closely related to common situation awareness. In short, to minimise confusion and ensure efficient use of assets all members of the COMAO must be aware of each other's roles and responsibilities. This is especially important when circumstances and/or roles change during a mission e.g. change of leader due to attrition. Furthermore, from the common data and information presented to the COMAO, each individual aircraft must have the ability to select and display information to the crew that is appropriate to its specific role and assigned task.

3.2 Improved reactive Capability

Reactive capability, that is the ability to respond to changed circumstances, is closely linked to situation awareness and relates directly to flexibility. In essence, it is the decision and action phases of Boyd's OODA cycle. The key to an improved reactive capability for future COMAO is, assuming adequate situation awareness:

- Accurate and timely situation assessment;
- Better informed decision making;
- More effective action.

Situation assessment and decision making

Whilst pre-briefed gameplans and Standard Operating Procedures (SOPs) are likely to remain an effective and coherent way of dealing with changed circumstances, they are by no means comprehensive. Against an unexpected SAM threat, for example, the mission commander must perform an accurate and timely situation assessment and decide the most appropriate course of action. This may involve re-routing the whole COMAO package around the threat (whilst remaining within a designated corridor) or aborting the mission entirely. In either case, a significant number of factors and the impact of particular courses of action need to be considered in a short space of time:

- Threat (numbers, type, capability);
- Re-route options;
- Fuel, time and profile implications;

Future COMAO will therefore require both situation assessment and electronic decision support systems to ensure both a high tempo of operation and an appropriate response to such circumstances. These studies have focused on improving the effectiveness of the aircrew/single aircraft combination through intelligent data processing, planning, and implementation functions to assist the aircrew. To extend electronic support tools to support multiple aircraft operations, we believe that a number of additional requirements should be taken into account. Including:

- hierarchy of command;
- synchronisation with external supporttools;
- different cooperative roles of group members;
- synergy.

The efficacy of such requirements in relation to future COMAO, however, remains to be proved. Inter-operability is likely to be a major issue in multi-national operations, especially with non-NATO forces. Different nations will undoubtedly possess different aircraft types, levels of technology and procedures. Aircraft equipped with electronic support systems may be able to compensate for this, but there will always be an overarching requirement for greater co-ordination and communication in such situations.

Effective action

Current procedures and technology limit the number of effective actions a COMAO can undertake in response to unplanned events. In short, structure (i.e. procedures, technology, organization and training) produces behaviour, and changing theses underlying structures can produce different patterns of behaviour.

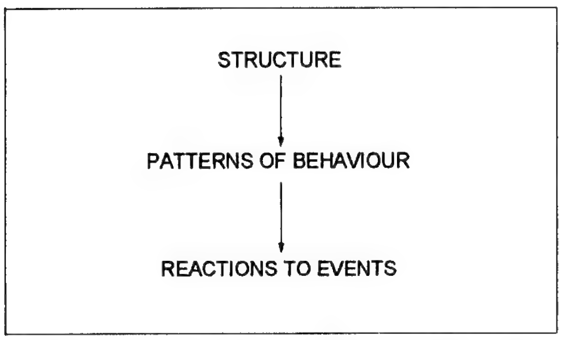


Figure 4: Reactions to Events - The Fundamental Explanation

By changing such structures more effective action can be realised, for example:

- In-flight Re-targeting:

  Should the main raid package suffer some attrition the highest priority Desired Mean Point of Impact (DMPI) for the raid can be re-assigned to the most appropriate aircraft;

- In-flight Re-roling:

  In completing one task, some aircraft may be able to re-role whilst airborne. F-15Es in Bosnia, for example, regularly undertook air defence missions (i.e. combat air patrols) following their ground attack mission, once they had refuelled from a tanker;

- Holistic and cohesive defence:

  An omni-directional 'defensive shield' against a variety of threats can be realised through a combination of new technology, tactics and training.

3.3 Improved Co-ordination

COMAO by definition requires co-ordination. However, this is merely a means to greater synergy within a group i.e. achieving an optimum combined effort. To exploit 'The Principles of War' and improve effectiveness future COMAO will require increases in both synergy and cohesion. A brief discussion follows:

Synergy

A group of co-operating elements is said to exhibit synergy when the capability of the group as a whole is greater than the sum of the capabilities of the individual elements. Synergy results from both diversity between elements (i.e. different roles, skills and attributes) and effective co-operation between those same elements (i.e. teamwork). Over the last fifteen years, extensive research has been undertaken into synergy and effective teamwork [Ref 5]. This research has identified a number of key characteristics which are required to formulate a 'team', including:

- Definable membership: the members must be identifiable, by name or type;
- Group consciousness: the members must think of themselves as a group;
- Sense of shared purpose: the members must have a common objective;
- Interdependence: the members need help from each other to achieve the common objective;
- Interaction: the members communicate with each other, influence each other, and react to one another;
- Unitary system: the members work together as a single entity.

In summary, we have concluded that synergy can be increased by improving the way aircraft work together as a team. This can be realised through:

- Technology to support cohesion and procedural responses;
- Development of appropriate tactics concurrently with technology, in order to fully exploit the technologies and guide their development. This will require:

  Operator involvement upfront in product development (i.e. Feasibility onwards);

  Multiple man-in-the-loop simulation;

  Operational Analysis combat modelling to test the robustness of new tactics;

- Development and introduction of appropriate training schedules.

Cohesion

Cohesion is 'the act of sticking together'. Fundamental to this concept is the need for effective communication and common situation awareness. Without both of these cohesion is likely to be low or very tenuous. Even with these two fundamentals present, cohesion may still be temporarily broken by:

- one or more members leaving the COMAO for some reason e.g. low fuel state;
- a new member joining the COMAO e.g. rendezvous with tanker or fighter escort;
- the group temporarily splitting and rejoining later .

For future COMAO, therefore, there is a requirement to generate and maintain cohesion (throughout a mission) by means of the following:

Effective communication

All elements of the COMAO must be able to communicate with each other as and when required, and where appropriate with higher command authorities and external agencies. Whilst such communications will need to be clear and concise to be effective, they must be neither detectable nor decipherable by enemy forces or other (unauthorised) third parties.

In addition to secure voice, data links will be required to convey information between different elements of the COMAO and with external agencies. They will be required to cover a wide range of operation (e.g. from 10km to beyond 80km) and a variety of bandwidths and data rates (e.g. from a few kilo bits per second to beyond one megabits per second).

Common Situation awareness

Each element of the COMAO must possess a common 'big picture' of the total situation, as well as the situation applicable to his own objective and circumstances.

Benefits of greater synergy and cohesion

Once these fundamentals are in place, the COMAO package can operate as a cohesive team, and can create and exploit synergy in novel ways. Group asset management, for example, allows the total assets (e.g. sensors, weapons, countermeasures) of the COMAO to be utilised in an optimum manner for the benefit of the group as a whole.

Group sensor management and data fusion, for example, could provide the following benefits:

- Omni-directional sensing of threats;
- Co-ordinated acquisition and targeting through optimum sensor coverage, update rates, and accuracy;
- Early resolution of multiple contacts and 'ghosts';
- Covert sensing strategies (e.g. bi-static radar, or LPI radar plus several IRSTs);
- Multiple redundant sensing.

Group weapons management could provide:

- Optimum allocation of firepower - vital in air combat to reduce inefficient over-allocation of weapons;
- In-flight weapon to target re-allocation i.e. the ability to upgrade the desired mean point of impact (DMPI) to a higher priority ground target;
- Guided weapons hand-off to other members within the package.

Group countermeasures management could provide:

- An omni-directional 'defensive shield' against a variety of threats;
- Co-ordinated threat deception e.g. blink jamming;
- Optimum use of countermeasures e.g. more efficient use of expendables.

In summary, by taking account of the location, capability and status of individual elements of the COMAO, and providing a holistic and co-ordinated response, group asset management will bring greater synergy and effectiveness to future COMAO.

4 SYSTEM CHARACTERISTICS

To realise the required operational effectiveness improvements for future group operations (as summarised in figure 1), the group of aircraft must be designed to operate as a whole system. Furthermore, in the wider context of a multi-national task force, this 'System' must operate as part of a much bigger 'System of Systems'. As a consequence, like any system, the group of aircraft must possess a number of key system level characteristics including high degrees of robustness, dependability and flexibility. These characteristics are briefly examined below.

Robustness

To be robust in their operation, highly integrated group assets must be insensitive to single element failure. Key contributing elements to the group include platforms, sensors and communications. In each case the system needs to be designed so that partial or full loss of any of these leads to a graceful degradation in performance. At all times the integrated group must perform at least as well as the 'non-integrated' group would have done using current practices and conventions.

Duplication or distribution may provide protection against the loss of key platforms or resources.

To compensate for localised sensor failure, degraded sensors would be screened out by robust data fusion and the desired coverage would be maintained by re-allocation of unaffected sensors by the sensor management function.

A robust communications management system would aim to minimise disruption and ensure optimal recovery when links are re-established.

Sensors and communications should be jam resistant and, where possible, should minimise the probability of provoking ECM systems through covert behaviour.

Dependability

Whereas robustness is concerned with continued, gracefully-degrading, viable operation in unfavourable conditions, dependability is concerned more with high availability for use. If a system is dependable, no likely fault state would be capable of causing total loss of capability.

Dependable designs avoid dependence on the availability of single vulnerable unit or equipment.

Flexibility

Flexibility is essential for a group operations system. The group may come into being shortly before a mission. It may exist for a few missions only and then be broken up and the

member platforms become part of other groups within which they must make an equally cohesive but not necessarily identical contribution.

Inter-operability with aircraft of other roles and with allies' aircraft is essential. Within a multinational task force, other group members and force assets such as tankers, surveillance and C^2 might be provided by other nations. Thus, varied levels of integration may exist within the group with links across national boundaries provided by international standard data links.

Other forms of flexibility are required: additions and losses to the group must be accommodated as well as changes in mission.

5 SYSTEM ARCHITECTURES

A fundamental feature of the group operation may be that command is devolved from the base to the group. In such circumstances, command responsibility would lie with a single group commander. However, the computing support required to enable the distribution of tasks and information in a co-ordinated way may lie on one or few platforms or they may be distributed among many or all platforms. These extremes of approach are discussed below and are referred to as the "group leader" and the "fully integrated" architectural models.

5.1 Architectural Models

Group Leader model

Within this model one platform acts as "group leader", and provides all the management and co-ordination processing. Platform data is fused on the platform that gathered it and broadcast across the group. Each aircraft may have knowledge of 'total picture' by fusing data from other platforms with its own data. However this is essential only for the group commander platform since the native data fusion capabilities would be optimised to provide the necessary situation for their role.

Individual platform assets such as sensors and defences would operate autonomously in the absence of instructions from the command platform. However, the command platform would monitor the state of the group and issue corrections when divergence occurred and would issue new tasks as they arose. It would also provide the link to the outside world and would be responsible for the group dealing with its base. A back up platform (perhaps several) would be required to 'shadow' the group commander platform to effect seamless takeover when necessary through attrition or other loss.

The advantage of such an implementation is that only the command platform and its back-ups would require extra capabilities. Other platforms could contribute to such a group without undergoing change.

Fully Integrated model

Within this model, no single platform carries out all the processing required for group functions. All group assets are managed as a 'whole system'. For instance, data fusion and asset management would take place in a distributed processing architecture across the group.

Such a processing architecture would be designed to ensure graceful degradation of the system when capabilities were lost due to attrition or malfunction. Its great advantage is its power to utilise emerging technologies to provide every member platform with up to date, fully optimal information. This in turn would promote greater group effectiveness.

Other options, between these two extremes, exist and need to be examined. In practice, the choice of architecture will depend on a trade-off of communications bandwidth and latency, system robustness, total processing load, development risk etc.

5.2 Whole/Group System technology requirements

Whichever system model is adopted, there are three interdependent fundamental supporting systems level technology requirements:

- Distributed data fusion.;
- Distributed mission processing;
- Covert communications.

Distributed Data Fusion

This technology enables fully optimal data fusion to be distributed across a tactical group so that all member platforms will have full simultaneous access to a common group picture (i.e. common situation awareness). Such a system would be capable of distributing identity processing (i.e. identification of a specific threat type) just as readily as position and motion estimation. In this way track identity conflicts would be avoided automatically because every platform sees the full picture, derived from all the available data.

Distributed Mission Processing

This technology allows mission processing tasks such as sensor and weapon allocation (and possibly even platform tasking in response to pop-up threats) to be distributed among the platforms. Given the same view of the world, participating elements should all reach the same conclusion as to who does what and when. This involves the employment of situation assessment, task generation and planning functions to arrive at a common plan for the group as a whole. Conflicts must be avoided both within the group and, at a higher level, between groups.

The ability to do distributed automatic mission processing depends on the ability to disseminate the group picture among the members of the group, with insignificant latency and divergence of content.

Covert Communications

A reliable communications system will be required to support distributed processing. The system should be secure, covert,

high data rate and low latency. Technologies will be required to support short range (line of sight up to a few kilometres), medium range (up to maximum line of sight) and long range (beyond line of sight).

5.3 Component Enabling Technologies

Key 'component' technologies which will be required to support future COMAO include:

- Unambiguous and timely covert communication;
- Distributed sensing (i.e. spatially separated across the COMAO package);
- Group asset management:
- Sensor management across the group;
- Countermeasures management across the group;
- Weapons management across the group;
- Data fusion;
- Data and Information Management;
- An intuitive and high quality CVI;
- Situation assessment support tools;
- Decision support tools;
- In-flight adaptive planning tools.

These technologies and how they satisfy the functional requirements for enhanced group operations are discussed below:

5.3.1 Achieving Situation Awareness/Assessment Requirements

High Situation Awareness

The aim of high situation awareness is to shorten and improve the quality of the COMAO's Intelligence (i.e. Observation and Orientation) gathering. This can be achieved through the following combination of technologies:

- Distributed sensing (i.e. spatially separated across the COMAO package);
- Sensor management across the group:
- Unambiguous and timely covert communication:
- Data fusion;
- An intuitive and high quality CVI.

Common Situation Awareness

To ensure each member of the COMAO has common data and information from which to form a collective perception of the total group situation, the following technologies will be required:

- Unambiguous and timely covert communication;
- Data fusion;
- Data and Information Management:

  Before fused data can form the basis for decision making, it must be organised with respect to the immediate concerns and roles of the decision making system. It must do this in ways which are appropriate not only for the human, but also for other decision support systems.

  Database handling technologies to support data and information management are under development, including associative memory systems.

- An intuitive and high quality CVI:

  In recent times, the volume of information available to modern combat aircraft has 'mushroomed'. Graphical display systems allow information to be presented to the crew as required, selectively and as appropriate to the role and task concerned [Ref 6, 7].

  It is vital that information presentation techniques are developed to optimise crew workload, and ensure common situation perception across the group. Technologies for information management and the manner of its presentation are under development.

5.3.2 Achieving an Improved Reactive Capability

To provide accurate and timely situation assessment, better informed decision making, and more effective action the following technologies will be required:

- Situation assessment support tools;
- Decision support tool;
- In-flight adaptive planning tools.

5.3.3 Achieving Improved Co-ordination

To improve synergy and cohesion within a COMAO the following technologies will be required:

- Unambiguous and timely covert communication;
- Common and high situation awareness (see 5.3.1);
- Group asset management (e.g. bi-static radar, co-ordinated countermeasures - blink jamming).

6 CONCLUSIONS

This paper has examined Group Operations (i.e. missions & composite air operations containing more than one aircraft of similar or different types). In particular, it has considered current practice, key issues and trends, and has identified future potential from both operational and technical viewpoints. From this analysis we have concluded the following:

(i) Functional requirements for future group operations can be grouped into three broad areas: improved situation awareness; improved reactive capability; and improved co-ordination.

(ii) Key system level technologies to meet such requirements include:

- Distributed (i.e. across the group) Data Fusion;
- Distributed Mission Processing to provide:

Group asset management (i.e. sensors, weapons, countermeasures);

Data and information management;

Situation assessment support tools;

Dedicated wide band covert communications;

Intuitive and high quality CVI.

(iii) In exploiting such technology the following benefits to future group operations will be accrued:

- Affordability - through more efficient use of assets and improved survivability;
- Lethality - through improvements in co-ordinated (in-flight) weapons management;
- Flexibility - via group asset management (i.e. sensors, weapons, countermeasures across the group);
- Availability - via multiple redundant systems (across the group) and co-ordination to mitigate system failures and attrition effects;
- Survivability - through co-ordinated sensor and countermeasure management providing a 'defensive shield' against a variety of threats.

(iv) To realise such benefits in FOAS timescales industry and the military will need to jointly pursue the following activities:

Evolve In-Service Capability

Whilst the technologies we have identified are realizable within FOAS timescales, they should not bring about a revolution in capability when they enter service. It is preferable that in-Service capability should evolve with the technology. For example, minor modifications and upgrades to current avionics could provide modest improvements to current COMAO (e.g. the addition of track quality and latency information to target/threat reports distributed via tactical datalinks). Furthermore, an evolutionary route will enable inter-operability concerns (e.g. operations with less sophisticated, non-NATO standard aircraft) to be resolved earlier during the design phase rather than later when FOAS is in service.

Concurrent and evolutionary tactics & technology development

Tactics need to be developed concurrently with the key technologies to fully exploit the technologies and guide their development. As such Service doctrine and operational evaluation units should be involved upfront in product development (i.e. Feasibility onwards) and evaluation. This goes beyond the traditional 'cockpit layout' involvement of aircrew and seeks to develop new and different techniques for using the technology. Such development activities will require a broad range of methods and tools including operational analysis combat modeling; real time multiple human-in-the-loop simulations; and weapon system concept design tools.

Technology Feasibility/Development Studies

The key technologies we have identified are largely immature. To assess their feasibility, cost and risk, further more detailed studies including Technology Demonstration Programmes (TDPs) are required.

7 ACKNOWLEDGEMENTS

This paper is based on a study undertaken by the Industrial Avionics Working Group (IAWG), composed of British Aerospace, GEC Marconi and Smiths Industries Aerospace and Defence Systems. The authors would like to pass on their thanks to the other study team members (Mike Everett, Isobel Watt and Charlie Hewitt) and Colin Noonan for their editorial contributions and assistance in preparing this paper.

Particular thanks to the following organisations for their valuable inputs and support:

Royal Air Force Air Warfare Centre
AIRCENT Tactical Leadership Programme
Future Offensive Air System Programme
Defence Evaluation and Research Agency
Franco-British Euro Air Group

8 REFERENCES

1 NATO office of information & press; C-M(91)88(Final) The Alliance's Strategic Concept; 1110 Brussels, Belgium. November 1991. 1215/91.

2 Woolley G.A. Gp. Capt.; Peace Support Operations (Parts 1 & 2); (unpublished Royal Air Force journal).

3 Royal Air Force Airpower Doctrine; AP3000; HMSO.

4 Finn C.J. Wg. Co.; Air Campaign Planning; (unpublished Royal Air Force journal).

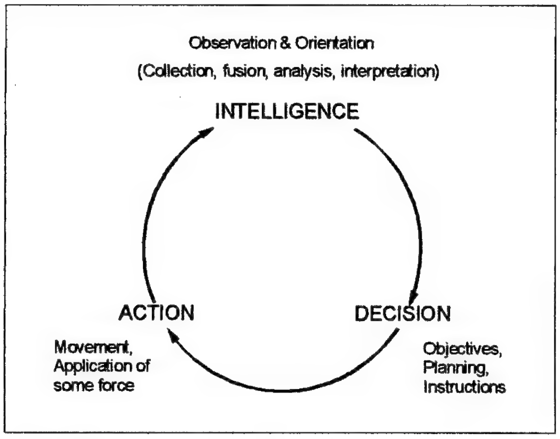
5 Swezey R.W, & Salas E: Teams: Their Training and Performance. ABLEX 1992. ISBN 0-893-91852-0.

6 Normanton T. H., Information Attributes and the Electronic Crewmember. Presented at the 4th Electronic Crewmember Conference, Kreuth, Germany. September 1997.

7 Ovenden C.R., Wykes K.M., Semple W.G., Normanton T.H., Aspects of the Crew Interface for Mission Systems. Presented at AGARD Paris 1997.

8 Hitchins D.K., Putting Systems to Work. Wiley 1992. ISBN 0-471-93426-7

Appendix A: The Boyd Intelligence-Decision-Action (IDA) cycle.



The Boyd Cycle - Intelligence, Decision, Action

Boyd's Observation, Orientation, Decision, Action (OODA) cycle is now often referred to as the Intelligence-Decision-Action (IDA) cycle. It applies to any 'system' capable of some form of cognition, making a decision based on acquired knowledge, and taking some form of action to implement that decision. Any such system (human, animal or machine) looking to achieve an objective will go through this cycle. The IDA cycle is described below and is summarised in the diagram above.

The Intelligence Phase

The cycle notionally begins with the Intelligence phase and is equivalent to Boyd's Observation and Orientation phases. It involves the following functions:

- Gathering information (usually involving some form of sensing)
- Processing of the information to arrive at a perception of the situation confronting the system
- An evaluation of how the perceived situation may evolve with time

Intelligence also includes the assessment of the system's own state and how it may change. The product of the Intelligence phase might be termed 'situational awareness'.

The Decision phase

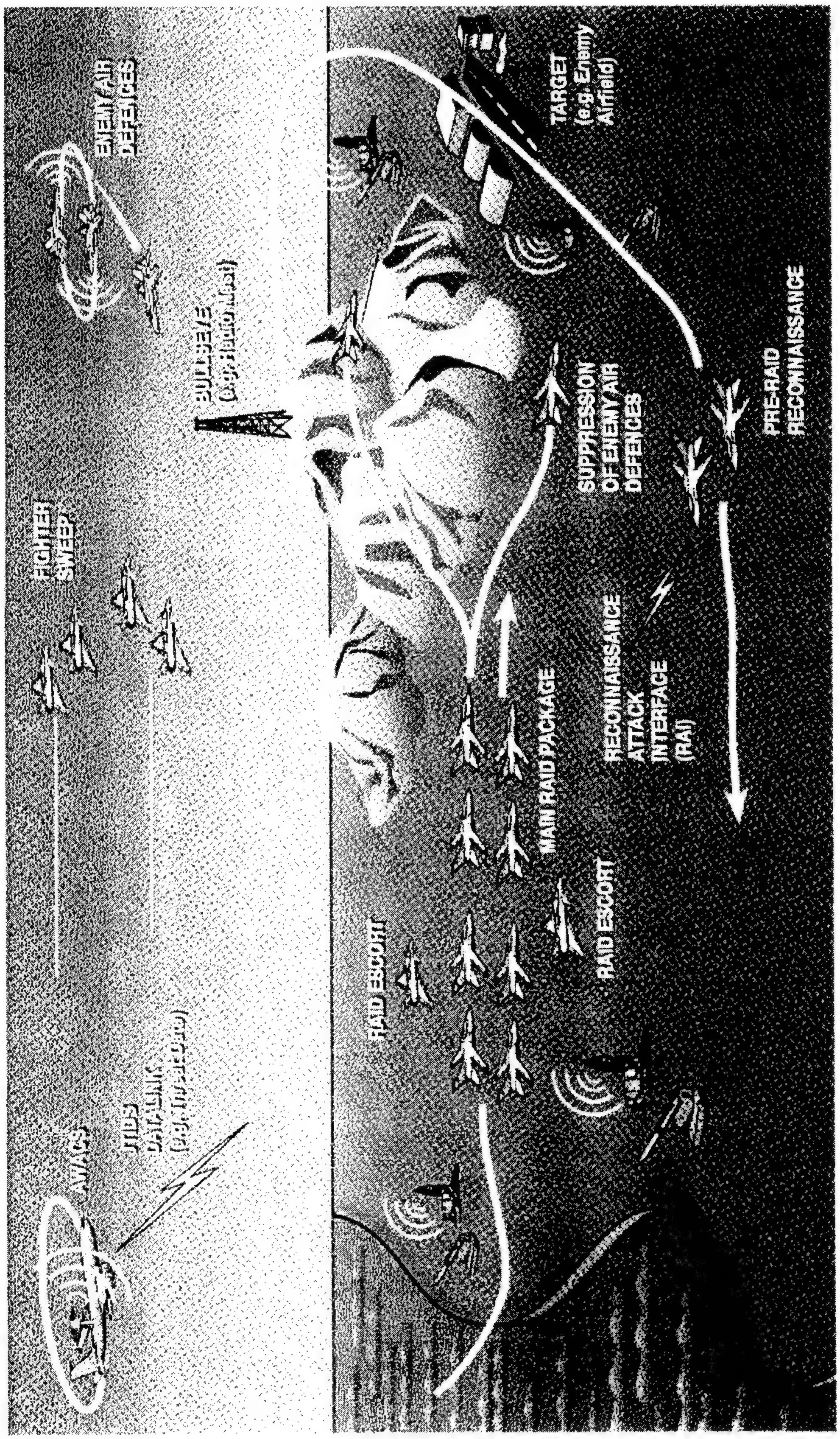
Based upon its perception of the situation confronting it and its own state, the system will decide on some aim or objectve. The system will then formulate a plan of action to achieve the objective.

The Action Phase

The plan of action is implemented, typically through a movement or an application of some form of force.

Note, that for most systems the element responsible for the Intelligence function will continue to operate while the Decision and Action phases are undertaken. Furthermore, in many instances a system will need (or choose) to achieve a number of 'intermediate' objectives in order to achieve its final aim. It will therefore go through a series of IDA cycles on its way to achieving the ultimate objective.

Appendix B: Figure 2: Typical Composite Air Operation



Collaborative Failure in Distributed Crew Systems

S.W.A. Dekker and C. Fairburn
Centre for Human Factors in Aviation
Linköping Institute of Technology
S-581 83 Linköping
Sweden

1. SUMMARY

This article describes the anatomy of collaborative failure in distributed crew systems. In a number of incidents and accidents across a variety of application domains, we have recognized situations that evolve towards failure only through a series of interactions between critical decision-makers and their wider operational environment. Various military operating systems show a growing reliance on distributed decision making through multi-agent communication and coordination. This means that *mis*coordination can translate smoothly into breakdowns in the entire system. Redundancy in the form of cross-checking and solliciting information from other system participantscan can produce a side-effect: if a system's success is distributed, so may its failure. In this paper we attempt to analyze some of the factors behind collaborative success and failure in distributed systems.

2. INTRODUCTION

In July 1994, flight 1016 crashed in heavy rain and windshear at Charlotte/Douglas airport in North Carolina. During the last stages of its flight, the DC-9 was confronted by a 40 knot headwind which changed into a 30 to 40 knot tailwind in fewer than 20 seconds. This was later classified as the most "severe" possible windshear as expressed by a NASA investigation. The shear was accompanied by a downdraft of 1800 feet per minute at its maximum, only seconds before impact (NTSB, 1995).

2.1. Collaborative failure

In its analysis of the flight crew's decision to continue an approach into severe convective activity that was conducive to a microburst, the safety board (NTSB, 1995) cites how lack of tight adherence to operating procedures during the approach into Charlotte led to and erosion of the crew's "situational awareness" (see pp. 100-106), including a degraded awareness of the weather situation around them. But closer inspection of the evidence surrounding this mishap actually reveals how the crew of 1016 actively explores the weather conditions on their approach path and final destination. Apart from continuously discussing among themselves the visibility and rain, and nursing their aircraft around towering clouds around Charlotte, the crew actively involve arrival controllers, tower controllers and other aircraft in the area and on the ground to build and up-date their assessment of the unfolding weather situation. This pattern of collaboration between air traffic control, weather services and pilots that appears necessary to have ultimately brought the mishap about is reflected more in post-accident litigation (where "blame" is shifted and re-distributed across the various participants) than in the official investigation.

This case of collaborative failure, where participants across the entire system cooperated successfully in steering 1016 into hazard, has echoes in a variety of other domains. The shootdown of the Libian airliner over Egypt, the first and only AH-64 Apache attack helicopter fratricide of the Gulf war (Travis, personal communication) and perhaps to a lesser extent the fratricide of two Black Hawk helicopters over Iraq and the shootdown of an Iranian airliner by the U.S.S. Vincennes (see e.g. Rochlin, 1997) all reveal a common pattern. The distributed system of practitioners on the ground and in the air (and at sea) gradually gets caught up and entrenched in an erroneous assessment of their respective evolving situations.

2.2 Cognitive fixation

These are cases of *cognitive fixation* (also labeled as "cognitive lock-up" or "cognitive hysteresis") - the failure to revise situation assessment as evidence comes in over time (De Keyser and Woods, 1993). Here, practitioners receive mixed cues about the state of the world. Many of the early cues they get are highly compelling, but misleading. For instance, the pilots of flight 1016 feel no turbulence (a necessary condition for windshear as far as they were trained) and hear from every pilot in front of them that the approach was "smooth" all the way to touchdown. In both cases these earlier and highly compelling cues are ultimately either wrong or obsolete (the storm had quickly caught up with 1016's flight path, leaving other aircraft in front unaffected, but only just). As pressure to take a decision mounts, less compelling cues start to emerge that point to the growing divergence between the crews' assessment of the situation and the state of the world. 1016 flies into a "wall of water" short of the runway at Charlotte. But these pieces of evidence are less compelling. Nothing is inherently convincing about heavy rain: it is neither a sufficient nor a necessary condition for microbursts.

But "cognitive fixation" as single process in one head of one practitioner underspecifies the richness and interactivity by which situation assessment is brought about and reinforced in typical operational worlds. It is critical to recognize that the situations such as the ones mentioned above evolve towards failure only through a series of interactions between the critical practitioners (the ones closest to the final

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

decision) and their environment. Erroneous situation assessment on part of the respective crews is created, sponsored and confirmed through coordination and double-checking with other participants in the operational environments (controllers, other pilots, battleground commanders, etc.). In some sense, these are highly distributed cases of cognitive fixation. The compelling but wrong or rapidly obsolete evidence is provided almost without exception by the distributed system around the pilots - the architecture that is constructed around them specifically to provide redundancy, to deliver information (on wind, on targets, etc.) from otherwise inaccessible sources.

3. REDUNDANCY REVISITED

3.1 Cross-checking and redundancy

To a great extent, the worlds in the above cases rely on cross-checking and redundancy for their operational success and safety. Arriving aircraft at a large airport such as Charlotte, for instance, are embedded in a rich network of multiple air traffic control centers, other aircraft in the vicinity, and various meteorological services, that all contribute to the creation of a coherent, up-to-date, and reliable picture of the weather situation. Similarly, the success of other attack aircraft is based heavily on its supporting infrastructure. On modern battle fields, multiple and independent communication channels link critical decision makers in the cockpit to battleground commanders, radar sites, ground forces, satellites, and other aircraft. These support systems are largely permanent structures - built into the operational environment.

Dependency on a larger infrastructure is increasingly prevalent in airforces around the world, whose aircraft's effectiveness and accuracy hinges in ever larger part on smooth cross-service coordination - whether their missions consist of payload delivery, medical evacuations or other tasks (Walters, personal communication). Successful missions require teamplay among a variety of agents who are distributed across the battle theatre. This requisite teamplay, or multi-agent coordination, is becoming the organizational and technological core of systems such as the Swedish Airforce 2000 (Hellstrand, 1997) and the Joint Tactical Information Distribution System, or JTIDS (Fielding, 1996).

But empirical data from Charlotte and similar incidents show that large dependency on distribution of information - and the subsequent coordination that is necessary to share it around the infrastructure - can shift a system's primary source of vulnerability. A growing reliance on distributed decision making through multi-agent communication and coordination means that *mis*coordination can translate smoothly into breakdowns in the entire system. Redundancy in the form of cross-checking and solliciting information from other system participantscan apparently produce a side-effect: *if a system's success is distributed, so may its failure.*

Of course, redundancy and double-checking are not supposed to contribute to outcome failure[1]. And normally they do not. It has been a long established maxim in system safety engineering that redundancy means reliability. *The* way to build a reliable system from unreliable parts is duplication of those parts, or "overcompleteness" as some say (Hoffer, 1989). This principle of "duplication as a substitute for perfect parts" pervades the engineering approach to nuclear power plants and other complex, safety-critical systems (Wildavsky, 1988). But the idea of redundancy is not limited to engineered systems. According to the large number of human reliability analysis methods available, human beings are quite unreliable system components that form weak links in safety-critical operations. Not surprisingly, then, the concept of redundancy has been extended to groups, or organizations of humans as well (Landau, 1969; Bendor, 1985; Lerner, 1986; Rochlin, 1989) and also to ensembles of humans and machines (see Hutchins, 1995; Billings, 1996 for discussions on redundancy in two-pilot cockpits and cockpits augmented with flight automation).

3.2 Miscommunication

Redundancy and information distribution may increase coordination and communication overheads in the conduct of a particular task. *Mis*communication in an otherwise sound redundant system could therefore be a likely source of vulnerability (and *is* in some cases, see Dekker & Woods, in press). But miscommunication does not explain Charlotte or the other mishaps. If anything about the communications is remarkable, it is that practitioners in fact take ample opportunity - especially vis-à-vis resource constraints and potential opportunity costs inherent to their operations - to communicate, clarify, and re-evaluate. Messages back and forth are well-understood and acted upon. Overal coordination in all of these accidents appears surprisingly smooth and professional. If the outcomes had been success (landing on runway 18R at Charlotte, for instance), participants and observers could have said that teamwork paid off; that double

[1] Redundancy has been associated with complex system failure before, but the sense in which differs substantially. Wildavsky (1988) and Perrow (1984) explain how redundant or "overcomplete" systems are both safer and less safe than systems consisting of simple, single components. Through the much higher number of components and interlinkages, redundancy (i.e. engineered duplication and overlap such as in a nuclear power plant) exacerbates the ways in which local anomalies and failures can propagate and multiply. The more redundancy, the more unpredictable and unexplainable the unfolding pattern of anomalies tends to be, where root sources of common mode failures are obscured and operators baffled by the system's unexpected and complex interactions.

checking worked well; that redundancy provided reliability.

3.3 Outcome success and failure

The fact that we can discriminate between coordination and miscoordination *only with access to outcome knowledge* suggests that the kind of miscoordination which leads to system breakdown is not just a substandard or deficient form of normative or good teamplay; that it is not just the lopside of canonical or good communication. Even in highly distributed system architectures, it appears that error and expertise stem from the same source and follow the same patterns of communication; that the same coordination governs the expression of success as well as failure. The features that make the difference between success and failure, then, must be more subtle as well as more subtantive. To begin to uncover them, we need to elucidate and go behind the label of 'redundancy' and explore the complexity of the cognitive behavior of the critical practitioners and their infrastructures.

4. CLARIFYING REDUNDANCY

4.1 Redundancy and heterogenity

Not all redundancy is the same, of course (see Lerner, 1986). For instance, a distinction is often made between *duplication* and *overlap* (Rochlin, 1989). When duplicated, the same function is carried out by multiple people or multiple organizational units or by a human and a machine. Duplication occurs in many engineered systems, for instance in multiple simultaneously running ignition systems in aircraft engines. In contrast, we speak of overlap when multiple different units merely have some functional areas in common, and can potentially cover for one another. Overlap occurs in the mammalian body, where for example bone marrow can take over the production of red blood cells in case the spleen is damaged or removed.

As these examples indicate, redundancy easily translates across engineering, organizational, medical and psychological domains. Redundancy is treated much as a singular category that can be used to aggregate instances of overcompleteness in various worlds without much corruption of its original meaning. Specific instances of back-ups or overlap or duplication are compared as equivalent attributes of people or organizations or engineered systems. But this assumed homogeneity of the concept of redundancy is misleading, and easily confuses the discussion of how it either bolsters or corrodes organizational reliability (see for example Sagan, 1993).

4.2 Redundancy as distributed process

Redundancy as empirical reality is heterogeneous, and as a label it is unspecific. It is quite a different thing across the various contexts in which we look at it. For example, the kind of redundancy observed in the coordination-rich Apache or Charlotte situations is characterized by a remarkable fluidity. The way in which assistance or back-up is recruited by those critical practitioners under pressure reveals a much more complicated and unstructured pattern of behavior than the labels of duplication or overlap could capture.

Redundancy in the cases of Charlotte and the Apache fratricide is not necessarily limited to fixed attributes of the organization. It is not just a baked-in ingredient of the permanent structures that envelop pilots in the conduct of their respective tasks. In many senses, this redundancy is not an attribute at all, but an intentional *process* of coordinating with other agents - both humans and machines - to assess a developing situation from different perspectives. It is driven by the contingencies and requirements of the circumstances. And as a process that unfolds itself over time and space, it is subject to cognitive demands, psychological resource constraints and strategic trade-offs that help shape its eventual outcome (see Woods, Johannesen, Cook, & Sarter, 1994).

5. BEHIND REDUNDANCY

5.1 Organized social complexity

Although the fact is too obvious, it still needs mentioning: one practitioner cannot be in all necessary places at the same time. The environments in which they work are typically too complex, widely spread and diffuse. Effective problem-solving requires a synthesis of perspectives from diverging geographical locations, of different specializations and equippage. To know about the local wind on final approach, commercial pilots consult air traffic control, who in turn consult various displays and meteorological services that are in turn interconnected with one another and with all kinds of technological artefacts that are hooked up to measuring equipment in various locations. The resulting architecture ranks high in "organized social complexity" (LaPorte *et al.*, 1975) which means it has a large *number* of distinguishable units involved in one operation and that these units have quite a bit of *differentiation* in their tasks and a high number of *interdependencies* or functional linkages between them. The circumstances surrounding the Apache fratricide similarly show high degrees of organized social complexity.

5.2 Informal networking

What sort of mix of perspectives and specializations is required for effective problem solving appears to hinge critically on the particularity of circumstances. Data from Charlotte and the Apache fratricide indicate that the process of creating redundancy by pooling multiple perspectives only partially follows established protocols through permanent organizational structures. As we can see in both cases, input from other system participants is recruited spontaneously under the pressure of rapidly

changing situational requirements. Who gets to participate in the architecture is primarily determined by who has access to relevant portions of the operational environment. Pilots who have just landed at Charlotte should be in a good position to report on the local weather during final approach - hence flight 1016 sollicits multiple pilot reports from these aircraft via the airport tower. In other words, a system's distributed architecture becomes partially fluid when pressurized by uncertainty or rapidly changing circumstances.

5.3 Contingency theory

This finding is consistent with central tennets of contingency theory, which holds that the combination of environmental unpredictability and dynamism forces practitioners to organize and coordinate themselves in decision-making structures that are not only decentralized but also highly organic (Mintzberg, 1979). The flexible, self-renewing and organic nature of such decision making structures closely mimmicks the changeability and ambiguity of the world in which they function. This is the most functional response to an environment which produces problems of unique and novel configurations; problems that are ill-structured and have no precedent (Orasanu & Connoly, 1993). Under such conditions, coordination can no longer be planned or structured *a priori*, but has to brought about by interaction *in situ*.

Rochlin (1989) observed this kind of "informal organizational networking" on US Navy aircraft carriers. They found how emerging problems in this highly unstable and unpredictable environment were tackled and pre-empted successfully by informal, spontaneously created networks of practitioners. Because of the need for various kinds of simultaneous expertise and novel combinations of specializations, the participants of these networks might be distributed widely throughout the ship and its naval command hierarchy. The informal networks observed were problem-specific and implemented at a local (ship) level. Significantly, they knew no recognition within permanent organizational structures or written protocols. After a problem was dealt with, the network would dissolve again, leaving hardly a trace upon the larger organization. This would have been true for Charlotte too, had it not turned into an accident. The multiple local coordinations with other pilots and weather services via the tower would have been neither unusual nor recorded in any organizational memory as something significant. Upon a successful landing of flight 1016, it would have vanished as largely irrelevant, and quickly superceded by landings of subsequent aircraft still on the approach path.

6. DISTRIBUTED COGNITION

6.1 The concept

To see how outcome failure can be a side-effect - an emergent property - of these partially spontaneous processes of redundancy creation, we need to go deeper into the cognitive content of the coordinations and exchanges that bring this redundancy about. One perspective that illuminates how knowledge is represented and propagated in multi-agent architectures is *distributed cognition* (Hollnagel & Woods, 1983; Norman, 1991; 1993; Woods, Johannesen, Cook & Sarter, 1994; Zhang & Norman, 1994; Hutchins, 1995; Zhang, 1997). Distributed cognition takes not an individual mind, but the entire socio-technical system as its primary unit of analysis. The idea is that cognition is "public and shared, distributed across agents, distributed between external artifacts and internal strategies, embedded in a larger context that partially governs the meanings that are made out of events" (Woods et al., 1994, p. 35). The distributed framework is explicitly cognitive, in that it deals with how information is represented and how representations are transformed and propagated through the architecture of cognitive agents and their artifacts (both human and machine) in the performance of tasks (Hollnagel & Woods, 1983; Hutchins, 1995).

This paradigm can help us trace representations of knowledge through the distributed cognitive architectures that surround the Apache helicopter and flight 1016 into Charlotte. We can then begin to see where these representations undergo critical transformations, where they lose or gain content, or how their propagation becomes pressurized through a variety of resource limitations (e.g., time shortage, or attentional resources) that operate on the cognitive agents within the distributed architecture. Below we sketch precisely that, and attempt to identify some of the factors that can help tip the balance in favor of outcome failure in a variety of operational settings.

6.2 Cognition is distributed, but not evenly

Although situation assessment in the cases mentioned is a joint product of the distributed cognitive system, there are still cognitive "pressure points" within the architecture. Some practitioners (we called them "critical practitioners" - pilots in the cockpit in many of these cases) carry more than others the burden to ammas and integrate shifting and uncertain data from the various corners of their environment. The likely diffusion of inputs they get as a result of their networking may also mean that discordant pieces of evidence are pushed down indiscriminately into the individual cockpit. Under the pressure of resource constraints and opportunity costs, practitioners at the sharp end must reconcile potentially diverging, ambiguous and uncertain cues about the state of the world. It has been established before that resulting workload saturation may exacerbate tendencies to "fixate" on a particular

hypothesis and its confirming evidence (DeKeyser & Woods, 1993)[2].

6.3 The interface as weak link in the distributed architecture

Interfaces between human and machines can become data-passing bottlenecks, particularly in highly dynamic situations where evidence about rapidly changing conditions appears over time and in pace with events in the outside world. In some sense, the interface represents a disjoint or a weak link in the line of communication across the distributed architecture. Here, representations of knowledge change their shape (and sometimes even their content, qualitatively and/or quantitatively). Looking back at the interface in hindsight, it can often be claimed that data (or *the* critical piece of data) was presented to the practitioner. But this does not mean that the data was necessarily operable in the context in which it was presented. Data availability is not the same as data operability - the latter depending crucially on how much cognitive work the operator needs to engage in to extract meaning from his interface *in situ*. The windshear accident at Charlotte provides a compelling example of the data operability problem: the tower controller was presented with a number of unconnected sets of digits which represented wind speed and direction at the various corners of the airport. The burden was on him to amass and integrate these digital data to form a meaningful and coherent picture of the potential of a windshear situation across the entire airfield. The rapid development of a microburst quickly outdoes a human's capacity to recognize the situation from such disintegrated and underspecified inputs.

6.4 Knowledge organization may not map onto situational requirements

The common pattern across the various mishaps mentioned shows that organic and decentralized decision making constitutes a common (and often effective) strategy to pool expertise and aggregate knowledge in an unpredictable and dynamic environment. But Charlotte and the Apache fratricide reveal that there are side-effects to informal organizational networking - that the complexity and dynamics of circumstances can create disjoints between local knowledge requirements (on part of the critical practitioners) and global knowledge organization.

[2] Alternative behavior that has been observed under such circumstances is the other extreme: "thematic vagabonding". Here, practitioners quickly and superficially sample every new piece of evidence that comes in. But attention becomes fragmented, situation assessment swiftly becomes incoherent, and pracitioners turn into inert actors and decision makers (Doerner, 1987).

Knowledge is organized across the distributed architecture in a certain way, and where parts and pieces of that knowledge are, changes over time. Organic networking is intended precisely to catch up with these changes - it traces sources of knowledge as they progress through the environment and learn about local conditions in places where the critical practitioner himself cannot be at that moment. Charlotte shows, however, that environmental conditions can sometimes change more rapidly than networking can provide access to the shifting centers of information. An aircraft that was a mere two minutes ahead of flight 1016 was left unaffected - the microburst simply had not hit yet. One type of insulation against this vulnerability of informal organizational networking may lie in a cognitive system's greater familiarity with the patterns of evidence that constitute highly dynamic threats. This would increase an architecture's sensitivity to the kinds of evidence a critical practitioner needs to make his decision (see the last point).

6.5 Procedural gaps and ambiguities in knowledge transfer

Decision processes in informal organizational networking routinely bypass permanent organizational structures because existing procedures for knowledge exchange are often underspecified or insufficiently applicable to particular or novel circumstances (Perrow, 1984; Rochlin, 1989; Woods, Johannesen, Cook, & Sarter, 1994). But some of the cases mentioned show that informal networking is still not immune to all aspects of formal proceduralization. Existing operating procedures and other written protocols can leave local practitioners uncertain about their role in the larger architecture and thus create unintended hiccups in the organic flow of information and dynamic re-organizations of knowledge.

In the Charlotte windshear accident, for instance, questions arose about why controllers did not provide adequate weather information to the pilots. It was found that the rules stated that "The provision of additional services [such as giving weather information] is not optional on the part of the controller, but rather is required when the work situation permits" (Controller Handbook, paragraph 2-2). A supposedly unconditional rule ("*not optional*") is superceded immediately by a conditional, or qualifying circumstance ("*when* the situation permits"). In this case, proceduralization recognizes the importance of the distributed architecture (clearly outlining the duties of certain agents within it). But it also reflects the complexity and ambiguities present in the typical operational scenario. The way in which these ambiguities play out and influence a particular practitioner's behavior *in situ* probably depends on the familiarity of this practitioner and his wider organization with the pattern of evidence. Application of existing operating procedures may become more appropriate

with better recognition of threatening collections of cues in the environment.

6.6 Authority gradients between agents in the architecture

The various cases show that control of attention is governed in large part by the interactions between critical practitioners and their distributed architecture. In other words, mutual coordination for assessment of the situation determines largely which cues are focused in on, and which cues are paid less attention to. In cases where the distributed cognitive system is subsumed in large part by military hierarchies, we see that authority gradients between practitioners (and especially between commanding practitioners and the so-called critical practitioners) can interact with how these coordinations are endowed with meaning by lower-ranking operators. When superiors confirm a piece of evidence as being correct (Feltovich, Spiro, & Coulson, 1993) or explicitly encourage their troops to carry out an order even in the face of ambiguity (as appeared to be the case in various fratricides), the cognitive fixation of critical practitioners may be reinforced. The role of authority gradients has been replicated in laboratory studies (Milgram, 1974) and has been recognized in other highly ambiguous field situations with distributed command centers as well (Hersh, 1970; Kelman & Lawrence, 1972).

7 CONCLUSION

7.1 The architecture's familiarity with the pattern of evidence

Success and failure in complex distributed systems appear to stem from the same source: the coordination between various participants in the system's architecture. Their coordinations determine in large part what kind of knowledge is activated, where attention is directed, and what information flows where and how. The incidents discussed confirm that there is only a loose coupling between the process of coordination and its outcome (Woods *et al.*, 1994). The factors that distinguish between outcome failure and success cannot easily be traced to components of the coordinative process, such as the quality of communications or to the time taken (or time available) to assess the situation from multiple angles. These issues are unable to explain the breakdown. Instead, the factors that push a process toward outcome failure appear to lie more within the distributed architecture itself, within the very organizational structure (its participants, its procedures, its interfaces) that has been set up (in part spontaneously) to deal successfully with a singular threat.

Closer inspection of these factors reveals that a distributed architecture's unfamiliarity with the pattern of evidence (including unfamiliarity with its highly misleading cues) appears to be a common ingredient in the cases where breakdown occurs. Therefore, repeated exposure to ambiguous configurations of cues may tend to encourage organizational learning (which is perhaps confirmed by those conflicts where fratricides occur largely in the beginning). As LaPorte & Consolini (1991, p. 23) suggest on a hopeful endnote:

"Those in the organizations carry on intensive efforts to know the physical and dynamic properties of their production technologies, and they go to considerable pains to buffer the effects of environmental surprises. In most regards, the organizations come close to a...well-buffered, well-understood technical core."

8. REFERENCES

Bendor, J. B. (1985). *Parallel systems: Redundancy in government.* Berkeley, CA: University of California Press.

Billings, C. E. (1996). *Aviation Automation: The search for a human-centered approach.* Hillsdale, N.J.: Erlbaum.

Doerner, D. (1987). On the difficulties people have in dealing with complexity. In J. Rasmussen, K. Duncan, & J. LePlat (Eds.). *New technology and human errors.* London: Wiley.

De Keyser, V., & Woods, D. D. (1993). Fixation errors: Failures to revise situation asessment in dynamic and risky systems. In A. G. Colombo and A. Saiz de Bustamante (Eds.), *System reliability assessment.* Amsterdam: Kluwer Academic, 231-251.

Feltovich, P. J., Spiro, R. J., and Coulson, R. (1993). Learning, teaching and testing for complex conceptual understanding. In N. Fredericksen, R. Meslevy, and I. Bejar (Eds.). Test theory for a new generation of tests., Hillsdale, N.J.; Lawrence Erlbaum.

Fielding, N. (1996). God's Eye View. *The Mail on Sunday*, June 30.

Hersh, S. (1970). *My Lai 4: A report on the massacre and its aftermath.* New York, NY: Random House.

Hellstrand (1997). Flygvapnet 2000: Ett systemflygvapen (Airforce 2000: A system airforce). *Flygvapnetnytt, 4*, pp.25-27

Hoffer, W. (1989). *Freefall.* New York, NY: St. Martin's Press.

Hollnagel, E., & Woods, D. D. (1983). Cognitive systems engineering: New wine in new bottles. *Int. J. Man-Machine Studies, 18*, 583-600.

Hutchins, E. (1995). How a cockpit remembers its speed. *Cognitive Science, 19*, 265-288.

Kelman, H., & Lawrence, L. (1972). Assignment of responsibility in the case of Lt. Calley: Preliminary report on a national survey. *Journal of Social Issues, 28*(1), 177-212.

Landau, M. (1969). Redundancy, rationality, and the problem of duplication and overlap. *Public Administration Review, 29(4)*, 346-358.

LaPorte, T. R. (1975). *Organized social complexity.* Princeton, N.J.: Princeton University Press.

LaPorte, T. R., & Consolini, P. M. (1991). Working in practice but not in theory: Theoretical challenges of 'high reliability organizations'. *Journal of Public Administration Research and Theory 1(1)*, 19-47.

Lerner, A. W. (1986). There is more than one way to be redundant. *Administration and Society, 18(3)*, 334-359.

Milgram, S. (1974). *Obedience to authority.* New York, NY: Harper & Row.

Mintzberg, H. (1979). *The structuring of organizations.* Englewood Cliffs, N.J.: Prentice Hall, Inc.

National Transportation Safety Board (1995). *Flight into terrain during missed approach (NTSB/AAR-95/03).* Washington, DC: NTSB

Norman, D. A. (1991). Cognitive science in the cockpit. *CSERIAC Gateway, 11*(2), 1-6.

Norman, D. A. (1993). *Things that make us smart.* Reading, MA: Addison-Wesley.

Perrow, C. (1984). *Normal accidents: Living with high-risk technologies.* New York, N.Y.: Basic Books

Orasanu, J. & Conolly, T. (1993). The reinvention of decision making. In *Decision making in action.* Ablex publishing co.

Rochlin, G. I. (1989). Informal organizational networking as a crisis-avoidance strategy: US naval flight operations as a case study. *Industrial crisis quarterly, 3*, 159-176.

Rochlin, G. I. (1997). *Trapped in the net: The unanticipated consequences of computerization.* Hillsdale, N.J.: Erlbaum.

Sagan, S. (1993). *The limits of safety: Organizations, accidents and nuclear weapons.* Princeton, N.J.: Princeton University Press.

Wildavsky, A. (1988). *Searching for safety.* New Brunswick, N.J.: Transaction Books.

Woods, D. D., Johannesen, L., Cook, R. I., & Sarter, N. B. (1994). *Behind human error: Cognitive systems, computers and hindsight.* Dayton, OH: CSERIAC

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*, 87-122.

Zhang, J. (1997). Distributed representation as a principle for the analysis of cockpit information displays. *The International Journal of Aviation Psychology, 7*(2), 105-121.

Keynote Address: Balancing Science and Practice in Conducting Team Training Research

Joan Hall Johnston, Ph.D.
Naval Air Warfare Center Training Systems Division
Science Technology Division - Code 4961
12350 Research Parkway
Orlando, FL 32826-3275
USA

Introduction

For more than a decade, the Naval Air Warfare Center Training Systems Division, Orlando has been conducting team training research and development sponsored by the Office of Naval Research. Most of the research has been conducted with aviation and surface ship teams. This address outlines five guidelines and recommendations on balancing science and practice in conducting team training research in order to address real world problems.

Background

A number of real world problems have been driving the need to improve the effectiveness and efficiency of Navy team training. First, as systems and missions become more complex, training and learning systems must keep pace. Second, the move toward fewer crew members, reduced life cycle costs and a smaller infrastructure demand that training investments be optimized in order to meet the performance challenges with fewer resources. Third, the global nature of future warfare requires training systems that are flexible enough to be useful anywhere, anytime, and on-demand. Many aspects of current team training systems are in need of attention. Such systems are costly: many require that actual equipment be employed for even routine practice. Most, if not all systems, are not available for shipboard use, and they do not, nor are they designed to take advantage of existing and emerging learning methodologies and technologies. Significant gaps exist where research is needed. For example, advanced research and development is needed to extend and apply intelligent tutoring, distance learning, embedded training, and performance support technology to the training problem. Successful demonstration and implementation of this technology will result in a number of positive changes, including: reduced crew size, increased quality of system maintenance and performance, better task performance (in terms of time, accuracy, quantity, safety), reduced life-cycle and training costs, and enhanced readiness.

For more than a decade, the Naval Air Warfare Center Training Systems Division has been conducting team training research in order to address these important R&D issues. In particular, the Tactical Decision Making Under Stress (TADMUS) and Aircrew Coordination Training (ACT) programs have resulted in significant

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

experiences from which guidelines and recommendations can be derived for balancing science and practice in implementing team training research. For further reading, a set of references are provided at the end of this paper from which these guidelines were developed.

(1) Relative costs and benefits must be considered in conducting team training research. By its nature team training can be a very expensive endeavor, however, when certain such "team-dependent" factors as mission readiness, ship readiness, and safety are at stake, the expected payoff of team training is quite high. Therefore, the need for conducting team training research should be balanced against the reality of what an organization needs and what it can afford to do. Given that the dollar investment in team training is high, then the research goals should be practical which means the application of a sound scientific and theoretical approach to designing training should be given the highest importance. For example, the following training strategies conducted under the TADMUS program have resulted in significant improvements in team performance: (a) team adaptation and coordination training, (b) cross training, (c) team self correction training , (d) leadership training, (e) critical thinking training, (f) embedded team training, and (g) team dimensional training for instructors. To design the training research, TADMUS researchers drew from theories and research on shared mental models, analogical and case-based reasoning, naturalistic decision making, teams as information processors, and social cognition.

(2) Investing in and planning for team training research requires that such practical issues as choice of team composition, team tasks, and research test beds be considered. The more realism that can be reproduced in the research environment, the better chance the research findings can be generalized and applied to the actual environment. This is especially important when the task environment is highly dynamic and human performance is significantly affected. For example, if the team is required to process large amounts of data, the content and context of the information changes from moment-to-moment, and the consequences of failure (i.e., when lives and property are affected) are catastrophic, then this guideline has strong application. For example, the TADMUS program took great care in developing a realistic team task environment because application of the research results was imperative for achieving program goals (i.e., a response to the Vincennes Incident). The research program required conducting experiments with five-person air warfare teams using a PC-based network of computers that simulated a portion of a ship's combat information center. Event-based air warfare scenarios were devised by subject matter experts in order to realistically manipulate typical combat stressors (workload, ambiguity, and auditory overload). Because careful consideration was given to these crucial aspects of research method application and transition of TADMUS training products have been considerable since the start of the program in 1990.

(3) Establish credibility as soon as you can with the customer because it is a crucial factor in achieving research goals: understand the task/problem, employ subject matter experts, get user involvement, and create partnerships and ownership. Because the customer, the ultimate end user of the training products, is highly knowledgeable about the complexity and difficulty of the team task, they need to feel confident that you understand it, and respect their knowledge and expertise. Such respect will pay off in the long run because you want the customer to implement your hard work.

The first task is to develop an understanding of the team task and/or problem. Investment in a steep learning curve is inevitable because complex team environments require a lot of time to learn. The second task requires enlisting user involvement throughout the research process. Thirdly, create partnerships with program stakeholders so they benefit immediately from interim research products and reports. Use user language when reporting findings to them so they understand the rationale for using a particular procedure. Finally, get stakeholders to develop ownership in the product by enlisting them as having responsibility for part of the research implementation plan. We followed these guidelines under the TADMUS program and this approach has been effective in helping to develop follow on advanced technology research programs. Be aware to avoid giving away too much too soon; managing user involvement is key to meeting your research goals. . It is easy to try to please everyone all the time to avoid arguments and disagreements, but sometimes the user can hinder conducting valuable research because the research goals and findings may seem counterintuitive to the user. Take care to explain how the outcome of the research will ultimately benefit and result in cost savings.

(4) Educate, Educate, Educate. Team training research can take a long time to come to fruition. Therefore, it becomes necessary to educate the customer/user with ongoing results, and provide products that become available in the short term. Creating a vision of expected outcomes with demonstrations can help the user understand what their future payoffs will be while they are waiting for the final product. For the TADMUS program, we created a canned, but live team demonstration with an in house version of our team test bed. We varied the demonstration depending on whether we had knowledgeable fleet representatives or less knowledgeable visitors. We replaced students with credible "operators" when necessary. In addition, we designed numerous briefings and developed several videos to communicate our vision. Multimedia was especially useful in demonstrating the impact of our work in the real combat team environment.

(5) Collect Data—As Fast As You Can! And Provide Tangible Transitions. As time goes on the education process should include more and more data to back up your findings. There is nothing more frustrating than to communicate a vision of what you expect, but not have the data to back it up. With team training research this can be especially hard because it can take more than

two years to develop the research test bed, get subject matter experts to create scenarios, and develop performance measures before the data collection begins. Data collection should begin as soon as human performance can be evaluated using the testbed. Even measuring a subset of the team's interactions, not just perceptions, can serve to get the ball rolling in terms of reporting findings.

Summary

In this address we described five guidelines and lessons learned from over 10 years of balancing science and practice in team training research. Brief examples from the TADMUS program were provided to illustrate our guidelines. We hope that these lessons can help other researchers in guiding them in the complicated task of team training research. Below is a list of references that supported the development of this paper.

References

Cannon-Bowers, J.A., Salas, E., & Grossman, J.D. (1991). *Improving tactical decision makingunder stress: Research directions and applied implications.* Paper presented at the International Applied Military Psychology Symposium, Stockholm, Sweden.

Duncan, P. C., Rouse, W. B., Johnston, J. H., Cannon-Bowers, J. A., Salas, E., & Burns, J. J. (1996). Training teams working in complex systems: A mental model-based approach. In W. B. Rouse (Ed.), *Human/technology interaction in complex systems* (Vol. 8, pp. 173-231). Greenwich, CT: JAI Press.

Hall, J. K., Dwyer, D. J., Cannon-Bowers, J. A., Salas, E., & Volpe, C. E. (1993). Toward assessing team tactical decision making under stress: The development of a methodology for structuring team training scenarios. *Proceedings of the 15th Annual Interservice/Industry Training Systems Conference* (pp. 87-98), Washington, DC.

Brannick, M.T., Salas, E., & Prince, C. (Eds.). *Team performance assessment and measurement: Theory, methods, and applications.* Mahwah, NJ: LEA.

Kirlik, A., Rothrock, L., Walker, N., & Fisk, A. D. (1996). Simple strategies or simple tasks? Dynamic decision making in "complex" worlds. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 184-188), Santa Monica, CA: Human Factors and Ergonomics Society. Naval Air Warfare Center Training Systems Division (October, 1997). *Research and Development Project Summaries.* Orlando, FL: NAWCTSD.

Salas, E., Cannon-Bowers, J. A., & Johnston, J. H. (1997). How can you turn a team of experts into an expert team?: Emerging training strategies. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 359-370). Hillsdale, NJ: LEA.

Smith-Jentsch, K.A. Johnston, J.H., Cannon-Bowers, J.A., & Salas, E. (1997). Team Dimensional Training: A methodology for enhanced shipboard training [CD-ROM]. *Proceedings of the 19th annual meeting of the Interservice/Industry Training, Simulation and Education Conference*, Orlando, FL.

Smith-Jentsh, K.A., Payne, S. C, & Johnston, J.H. (1996). Guided team self-correction: A methodology for enhancing experiential team training. In K.A., Smith-Jentsch (Chair), *When, how, and why does practice make perfect?* Paper presented at the 11th annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Swezey, R.W., & Salas, E. (Eds.) (1992). *Teams: Their training and performance*. Norwood, NJ: ABLEX.

Zachary, W., Bilazarian, P., Burns, J., & Cannon-Bowers, J. A. (1997). Advanced embedded training concepts for shipboard systems [CD-ROM]. *Proceedings of the 19th annual Interservice/Industry Training, Simulation and Education Conference*, Orlando, FL, 670-679.

The Effects of Stress on Individual and Group Problem Solving

J. D. Callister, Ph.D.
Maj, USAF, BSC
DoD Medical Support
AL/AOCN, 2507 Kennedy Circle
Brooks AFB, TX 78235-5117, USA

G. L. Percival, Ph.D.
Capt, USAF, BSC
USAF Survival School
Fairchild AFB, WA 99011, USA

P. D. Retzlaff, Ph.D.
Psychology Department
University of Northern Colorado
Greeley, CO 80639, USA

R. E. King, Psy.D.
Maj, USAF, BSC
USAF Research Laboratory
Wright-Patterson AFB, OH, USA

Summary

The United States Air Force (USAF) Survival, Evasion, Resistance, and Escape (SERE) training offers an ideal setting for studying the effects of realistic stress on individual and group problem solving. In the past, the effects of stress on individual and team performance have been observed and modified to enhance training. Currently, the effects of stress are being studied by collecting objective data, and systematically controlling program modifications to enable data-based decisions. Initial data suggest that cognitive performance, including problem solving and associative learning declines during training. Student's self-report of fatigue and competence to perform necessary individual and group skills also changed considerably during training.

Introduction

Many factors, including stress, affect the quality and efficiency of group problem solving (Stokes & Kite, 1994). Understanding these factors and finding ways to counteract their effects has direct operational relevance. Much of the research in this is area has been of limited utility because stress research conducted in laboratory settings is often extremely artificial and the experimental conditions can not be considered "stressful" in a practical sense (Flor & Turk, 1989). Recently, significant improvements have come about with the use of operational teams using high fidelity simulators. Similarly, the USAF Survival, Evasion, Resistance, and Escape (SERE) training programs offer an ideal circumstance to study the effects of stress on individual and group problem solving (Callister & Percival, 1997).

During a 17-day program, SERE students are given didactic and experiential training. Throughout the course, students are required to solve a variety of individual and group problems. These problems are developed for training, but are based on real events. The instructor cadre continually works to enhance the realism of the training environment incorporating the most recent intelligence. This training environment has many advantages for studying human performance, but the two principal advantages are that all aspects of training are closely monitored, and many aspects of training can be experimentally manipulated while remaining ecologically valid.

Many of the well-known effects of stress on individual performance have been observed, such as degradation resulting from sleep deprivation, fatigue, lack of nutrition, disorientation, surprise, intense emotions, and loss of control. Recently, SERE instructors raised concerns about their students' ability to benefit from training given their observed cognitive performance decline resulting from sleep deprivation. These instructors asked that cognitive performance parameters be measured. Also of interest is the relationship between cognitive performance changes and subjective reports, and their relationship to individual and group problem solving, since SERE training culminates with a series of multi-group problems to be solved. Various elements are put in place to hinder situational awareness, communication, command and control. Students are highly motivated to solve the problems since they believe that the sooner they solve the problems the sooner they will be released. Instructors' anecdotal reports suggest a point at which group problem solving completely disintegrates, and it is our aim to identify the specific factors that lead to these failures. It is our hope that such study will lead to a better understanding of the effects of stress on individual and group problem solving, and more importantly, will point to valid countermeasures which can be taught to military operators.

Method

Subjects. All USAF aircrew are required to complete SERE training. This population includes pilots and navigators as well as enlisted aircrew and non-aircrew such as loadmasters, pararescue, and combat controllers. About 90% of SERE students are male and most are between the ages of 18 and 29.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

This paper describes data collected from an initial sample of 15 subjects. There was one female student and five officers in this sample. The mean age was 24.9 (SD = 1.97).

Measures. Several psychological tests and measures were used to assess cognitive functioning, fatigue, self-assessment and problem solving in SERE students. The Spaceflight Cognitive Assessment Test (S-CAT) was used to measure changes in attention/concentration, memory, pattern recognition, and problem solving. The Sustained Operations

Assessment Profile (SOAP) was used to measure changes in self-report of cognitive, affective, and arousal dimensions of fatigue. A self-assessment instrument was used to measure student's perceptions of their competence to perform specific skills taught during SERE training.

The S-CAT was developed by NASA to monitor functional changes in shuttle crewmembers during space missions. The S-CAT includes four brief cognitive tasks and results in six scores. This test is administered on a laptop computer and requires approximately 12 minutes to administer. The six variables measured by the S-CAT are Symbol Recall (SRA); Vigilance: Accuracy (VA) and Speed (VRT); Math: Accuracy (MA) and Speed (MRT); and Pattern Matching (PMA). The symbol recall task requires the subject to learn matched pairs of numbers and symbols. The vigilance task requires the subject to monitor letters presented on the computer screen for two minutes. The math task requires the student to solve 20 simple addition and subtraction problems. The pattern-matching task requires the subject to remember non-verbal patterns, measuring graphic/ spatial memory. Both accuracy and reaction time are measured for Math and Vigilance tasks.

The SOAP is a self-report measure of cognitive, affective, and arousal dimensions of fatigue. The SOAP was developed to measure fatigue in military and civilian personnel engaged in sustained operations (Retzlaff, King, Marsh, and French, 1997). Ninety items are administered with the subject responding on a 1 to 5 scale. The test is administered in paper-and-pencil format and requires approximately 4 minutes to administer. The scales of the test include three cognitive dimensions (Poor Concentration, Boredom, and Slowed Reactions), three affective dimensions (Anxiety, Depression, and Irritability), and four arousal dimensions (Fatigue/ Low Energy, Poor Sleep, Work Frustration, and Physical Discomfort).

The self-assessment questionnaire of competence to perform SERE skills was developed by the SERE staff to evaluate student's perceptions of the quality of training. Students are asked to rate their confidence to perform a variety of individual and group skills such as to "take command if senior," "aid others to escape," and "bounce back and continue resisting," among many others. Ratings are made on a 5-point scale, with 1 representing disagreement and 5 representing agreement with one of thirteen statements. This measure takes a few of minutes to complete.

Design. Cognitive performance and fatigue measures were administered to each participant six (6) times during the 17 days of training. The schedule of test administrations can be seen in Table 1.

The self-assessment questionnaire was administered three times during training; first, during the initial orientation, day 1; second, during resistance lab exposure, day 11; and third, during the last academic training before graduation, day 17.

Table 1. Test Schedule

Test	Day	Training	Comment
1	1	Academic Training	Baseline
2	5	Academic Training	Baseline
3	11	Resistance Exposure	Physical & Emotional Stress
4	15	Resistance Training	Following Three Days Rest
5	16	Resistance Training	High Stress, Problem Solving
6	17	Academic Training	Relief from Stress, 18 hrs Rest

SERE training begins with five days of academic training in survival and evasion skills. All academic training consists of classroom instruction, and students have access to high quality food and sleep conditions. Following academics, students spend six days in the field practicing survival and evasion skills. During this field training, students participate in strenuous physical activities, sleep is restricted and degraded due to environmental conditions and training requirements, and nutrition is severely restricted. On the last day of survival and evasion training, students are introduced to resistance training. During this exposure, significant cognitive and emotional demands are placed upon the students on top of the physical stress of field training. Upon release from resistance exposure, students receive three days of academic training in resistance skills. During this period they are allowed time to recover from the previous stress. They have ready access to high quality food and sleep conditions. Following resistance and escape academics, students spend two days practicing their new skills under realistic conditions. During this phase students face significant physical, emotional and cognitive challenges. Opportunities for sleep and nutrition are severely restricted. Upon completion of this phase of training, students are given an opportunity to eat and sleep, and then they complete a half-day of academic review before graduation.

Results

Cognitive Performance. In general, cognitive performance varied during training. Figure 1 represents mean accuracy scores for Symbol Recall Accuracy (SRA), Math Accuracy (MA) and Vigilance Accuracy (VA). All three accuracy scores declined during the survival portion of training and then improved with rest between the survival and resistance portions of training. Symbol Recall declined significantly during resistance training and continued to decline through completion of training. Pattern Matching Accuracy, not shown in the chart, actually increased during the survival portion of training.

Figure 1. S-CAT Accuracy Variables

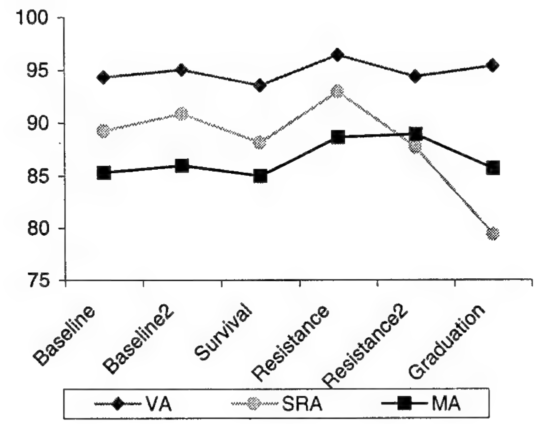
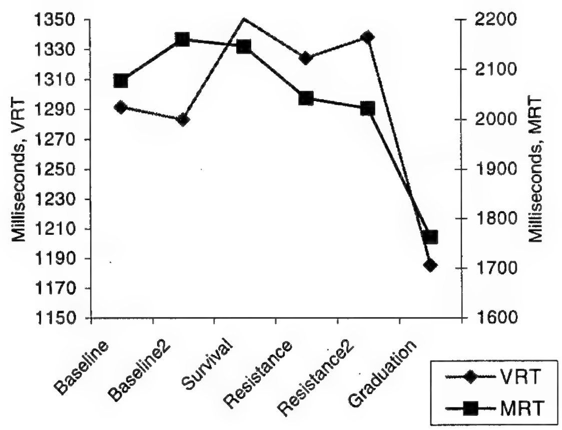


Figure 2 represents mean response times measured in milliseconds for Vigilance Speed (VRT) and Math Speed (MRT). Response time for Vigilance increased significantly during survival and resistance training, and then decreased dramatically by the end of training. Response time for Math decreased throughout training.

Fatigue. Self-report of fatigue varied dramatically during training. All dimensions of fatigue increased during the survival and resistance portions of training. Most dimensions also showed marked improvement with rest between these portions of training. The only exceptions to this were Anxiety and Poor Sleep that did not improve until the completion of training. A few examples of mean scores on SOAP variables can be seen in Figure 3.

Figure 2. S-CAT Speed Variables



Self-assessment of SERE Skills. Student's ratings of their confidence to: a) aid others to escape, b) take command if I am senior, c) bounce back and continue resisting, etc. changed during training. Figure 4 represents mean ratings across training. High initial ratings drop following survival and exposure to resistance training. Ratings then rise significantly by the completion of training.

Figure 3. Selected SOAP Variables

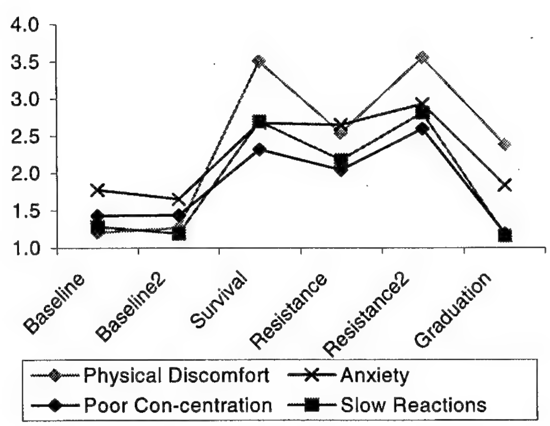
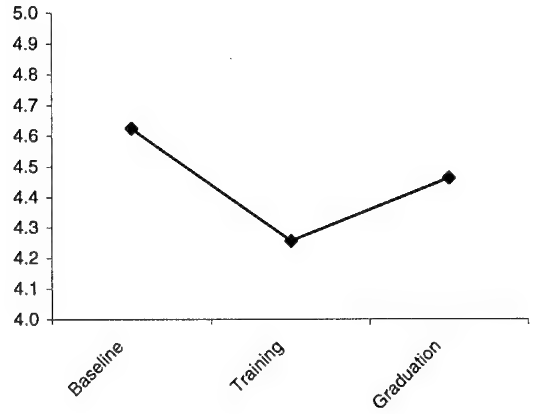


Figure 4. Competence to Perform SERE Skills



Discussion

The stress of SERE training is evident in the changes in cognitive performance, fatigue, and competence during training. While spatial memory changed very little, problem solving, associative learning, and vigilance changed dramatically during training. Students' ability to attend to environmental cues and solve simple problems decreased with the stress of training and improved with rest. Student's ability to learn new associations was even more sensitive to the stresses of training and continued to worsen through the completion of training.

These changes in cognitive performance correspond well with students' reports of increased fatigue, across cognitive, emotional and arousal dimensions. Self-report of fatigue seems to be very sensitive to the effects of training demands, with clear improvement in most aspects of fatigue seen during the respite between the survival and resistance portions of training, and at the completion of training.

Students' self-confidence appears to follow a consistent pattern. Initially, students may be overly confident in their ability to perform SERE skills, but that confidence falls after realistic exposure to the task. By the end of training, students appear to have gained what may be a more realistic sense of competence to perform the individual and group skills required of them in a SERE situation.

The results of this study have significant implications for those responsible for training. A realistic training environment is important, but the exposure to the stress of a realistic environment directly affects student's abilities to learn. Therefore, realism and learning ability must be balanced to provide the best training environment.

References

Callister, J. D. & Percival, G. L. (1997, August). *The effects of sleep deprivation on cognitive performance.* Paper presented at the DOD Joint Services Survival, Escape, Resistance, and Escape (SERE) Agency Psychology Conference, U. S. Air Force Academy, CO.

Flor, H. & Turk, D. C. (1989). Psychophysiology of chronic pain: Do chronic pain patients exhibit symptom-specific psychophysiological responses? *Psychological Bulletin, 2,* 215-259.

Retzlaff, P. D., King, R. E., Marsh, R. W., & French, J. (1997). *The Development of the Sustained Operations* TR *Assessment Profile (SOAP).* (AL/AO-TR-1997-094). Brooks AFB: Armstrong Laboratory.

Stokes, A. & Kite, K. (1994). *Flight Stress: Stress, Fatigue, and Performance in Aviation.* Brookfield, VT:Ashgate.

The Effects of Sleeploss, Continuous Performance, and Feedback on HierarchicalTeam Decision Making

Robert P. Mahan
Department of Psychology
The University of Georgia
Athens, GA. 30602
U.S.A.

Linda R. Elliott
Air Force Research Laboratory
Brooks AFB, Texas

Philip T. Dunwoody
Christopher J. Marino
The University of Georgia, Athens, Georgia

SUMMARY

Advances in the performance of command and control activities in global and tactical military operations will need to address the issues associated with the interaction of team level decision variables with operational readiness constructs. For example, team decision making is currently a significant area of research and training, yet very little applied significance has been attached to the effects of operational factors such as sleeploss and continuous performance on team-level constructs. This study explores the effects of sleep loss, continuous performance, and the presence and absence of task performance feedback on team decision making.. The study demonstrates that these operational variables are associated with significant performance decrements at a number of levels associated with the team hierarchical decision making model. The implications of these performance decrements for complex team performance are briefly discussed.

1. INTRODUCTION

The majority of team decision making research paradigms offer only snapshots of the complex nature of team behavior. Worse yet, implicit in many of the research interpretations based on these snapshots of behavior is the notion that these outcomes can be extrapolated (generalized) to real work settings where teams operate under a variety of hostile conditions, such as reduced sleep and elevated levels of fatigue. For example, it is well known that an important constraint on job performance in military, as well as many non military work environments, is the demand for sustained and continuous work, which is often performed during periods of inadequate rest (Ref 1; Ref 2). A fundamental component of operational readiness in military doctrine and training is preparing soldiers for sustained and continuous work that extends beyond normal duty hours. Command and control, medical, security, communications, navigation, and most transportation activities demand continuous and often sustained work. As a result of the

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

requirement for sustained and continuous vigilance, individuals and teams do not have adequate opportunities for rest, making sleeploss and fatigue highly influential factors in mission outcomes. It is noteworthy that sustained performance variables have been identified as being associated with the underlying causes of many industrial and military accidents (Ref 3).

The present investigation represents an on-going exploratory effort aimed at more in-depth studies on the cognitive effects of stress in the context of team decision making. The guiding premise of the present study is that teams are composed of different kinds of individuals with different types of expertise who work together to solve complex problems under very demanding and difficult circumstances. Task efforts require the combination of expertise from multiple team members in order to achieve particular operational objectives. In Figure 1, the performance process is represented as a hierarchical team problem where information is passed along to team members and terminates at a team leader who is responsible for a final assessment that leads to team level action. The objectives for this study lie in examining how continuous performance, sleeploss, and the presence versus absence of timely performance feedback combine to produce particular team-based behavioral and cognitive outcomes. The idea of team performance as a context in which to document the effects of general workplace stressors represents a step toward further development of a nomological network that connects stress phenomena to high-level team-based cognitive functioning.

2. METHOD

2.1. *Participants.*

Thirty-two male subjects divided into eight groups (teams) were paid participants for this pilot study. They were selected from undergraduate subject pool and were randomly assigned to the groups. They were paid six dollars per hour for the duration of their participation. In addition, each group (team) was given the opportunity to earn money on the basis of performance. The highest performing team received $160.00 to be split four ways, the second highest earned $80.00, the third-place team $40.00 and the fourth place team $20.00. Participants ranged in age from 22 to 26 with an average age of 23.4 years. Participants were screened for the display of stable sleep-wake behavior using a two-week log where they recorded details about when they slept, the duration of their sleep and nature of their work (Ref 4). All the participants in the study manifested stable diurnal sleep wake cycles and displayed sleep length values and social patterns that were similar to those found in permanent day shift industrial shift workers (Ref 5).

2.2. *Apparatus*

Four PC computers in a computer laboratory were used for the simulation platforms. A Novell Netware 3.12 network served the simulation software to the PC machines. The computer laboratory was closed during the study so that only study participants were allowed access.

3. PROCEDURE

3.1. *Training.*

The 32 subjects were randomly assigned to eight teams. Each team member was trained to achieve at 90 percent criterion accuracy rate in a series of four consecutive simulations prior to be classified as trained on the simulation. The training criterion for teams took an average of 3.2 hours to reach.

3.2. *Simulation.*

Each team of research subjects participated in a team networked simulation which was implemented with TIDE[2] (Team Interactive Decision Exercise for Teams Incorporating Distributed Expertise) (Ref 6). TIDE[2] is a software program that allows for simulating team-based decision making. Specifically, multiple information sources referencing a criterion state or object are presented to team members for evaluation. Team members are asked to render judgments about the criterion on the basis of the available information. The members learn to make these judgments through training. During training they learn to weight the information in forming their judgments in a manner that converges on a true criterion model that has been created by the researchers. Providing feedback to the team members after each of many judgments on the criterion does this. Feedback allows the team members to understand the causal relations between information sources and criterion values.

TIDE[2] is a flexible simulation that can be configured in a number of different ways to allow the study of different team-based decision making problems (Ref 6). However, the current application was programmed to simulate a Naval air defense operation involving four Naval platforms: Aircraft Carrier (Carrier), Advanced Warning Airborne Command System (AWACS), Aegis Cruiser (Cruiser), and Coastal Air Defense (CAD). These platforms represented the air defense component of a carrier battle group. Each member of a team learned to play the role of a particular platform in a four member hierarchical team where the role of the Carrier represented the team leader position. The other air defense platforms were operated by subordinate commanding officers, and thus configured the hierarchical structure of the decision making team. Figure 1 below summarizes the hierarchical team model showing distributed expertise components.

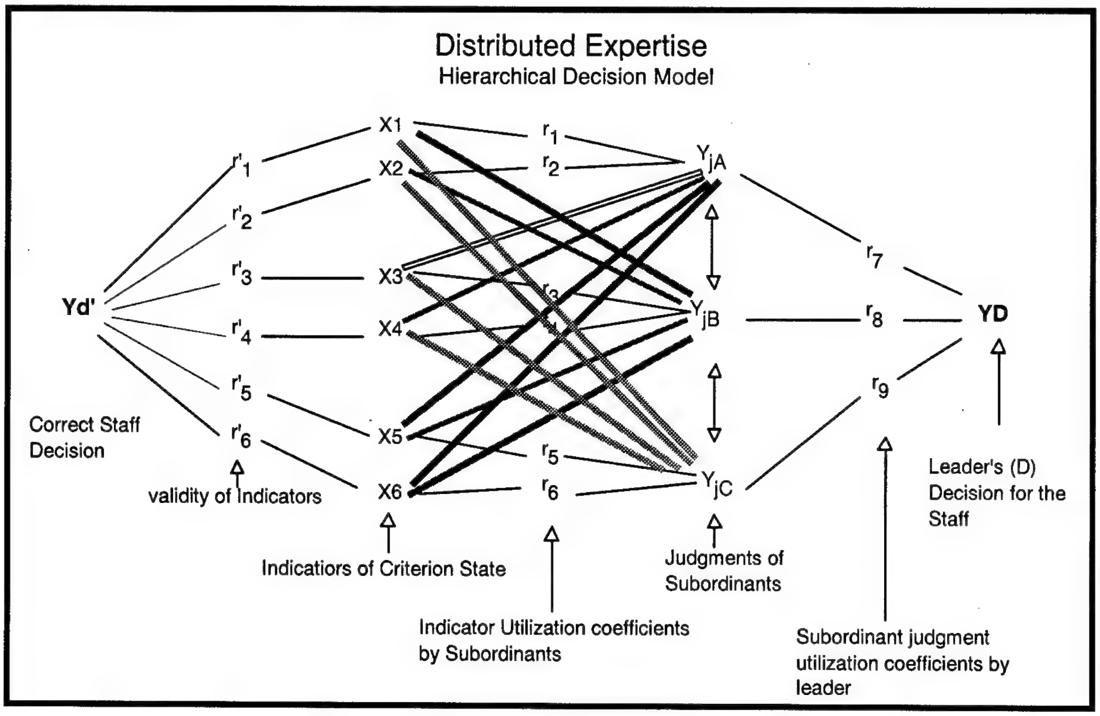


Figure 1. Hierarchical Decision Model

The task of the four member teams was to monitor the airspace surrounding the aircraft carrier. The eight teams performed this task for eight (8) consecutive hours under each of two manipulations- No Sleep Loss, and 24 hours without a sleep period (Sleep Loss); Outcome Feedback and No Outcome Feedback. The Feedback variable was used to simulate operational conditions where feedback is readily available to guide and support decision making, and during times when it is not available (i.e., the decision horizon is extended and requires many decisions before feedback on decision quality is known). The conditions were all counterbalanced.

3.3 *Task Overview.*

The goal of each team performing the air defense simulation was to monitor the airspace surrounding the Carrier and to evaluate targets that enter the air space on a number of dimensions (e.g., size, speed, angle, range). Teams then render a judgment about the action appropriate to the nature of threat the targets pose to the Carrier group. Judgments about threat were based on a 7 point scale made with regard to the aggressiveness of the action taken toward the target from "ignore" (lowest perceived threat- lowest aggressive action) to "defend" (highest perceived threat-highest aggressive action). Intermediate values were review (2), monitor (3), warn (4), ready (5), and lock-on (6).

Each member of the team was monitoring the airspace from his personal computer that was networked to the other stations. All judgments regarding a target were forward by subordinate team members to the Team Leader, which was the Carrier. A timer on each of the networked computers displayed the amount of time the team had in order to make the final decision on response action. Each decision required that the final decision be made within 4 minutes. Thus, teams were performing 15 decisions/hour. When the timer counters down to 30 seconds remaining each terminal began to beep, indicating the impending decision response limit. Once the team leader received the judgments from subordinate members, he made the final decision on action. No member of the team was privy to all the information necessary to successfully evaluate the target. Thus, members were required to query each other on the information they needed for their respective judgments and transmit the information requested to those queries (Ref 7). All team members made the same threat response judgments.

4. RESULTS

Figure 2 demonstrates that the ability of the team leaders to execute their learned judgment policies for integrating the tactical information was poorest in the No Sleep-No Feedback condition combination. This decrement manifested itself as a loss in the ability to control the execution of knowledge. Multiple regression of team leader judgments on information cues (Consistency) showed that over time, the team leaders became progressively worse at weighting and integrating the information, with the poorest performance occurring in the sixth and eighth hour epoch.

Dyadic interactions among team leaders and subordinate team members are particularly important in hierarchical distributed decision making, and it is clear from the data in this study that continuous work coupled with the absence of feedback during sleep loss affects team hierarchical decision making in complex ways. In some sense, the absence of both statistical and methodological power suggests that these effects are likely to be significantly underestimated in the present research. Typical military doctrine call on for at least a three-fold increase in the continuous performance

parameters of the type studied here. Thirty-six hour continuous tactical exercises are not uncommon in current military operations where the absence of forward basing requires the round-trip deployment of bombers and other aviation systems from U.S. mainland locations to distant threat sites.

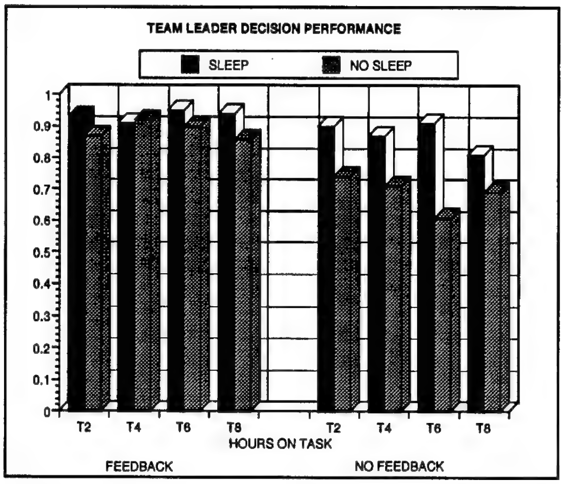


Figure 2. Team Leader Decision Performance

Figure 3 shows another clue as to the nature of the performance decrement in Team Leaders was revealed in the hierarchical sensitivity measure (HS). During training, Team Leaders learned to use subordinate judgments effectively in generating criterion estimates. This was done vis-a-vis leveraging outcome feedback. Thus, after many judgments of criterion values based upon subordinate judgments (Team Leader cues) with outcome feedback following, the Team leader was able to de-bias subordinates' judgments in generating criterion estimates that met the 90% training criterion. Thus, indexing the absolute difference between the least squares regression solution of criterion scores on subordinate judgments, and the Team Leader judgments on subordinate judgments (i.e., the cues), would give the hierarchical sensitivity to subordinate evaluations of the criterion. The figure below shows that this sensitivity was poorest during the No Sleep-No Feedback condition. On average and over time, the Team Leaders lost their ability to successfully de-bias the subordinate judgments in a manner that converged on the least squares "trained" solution.

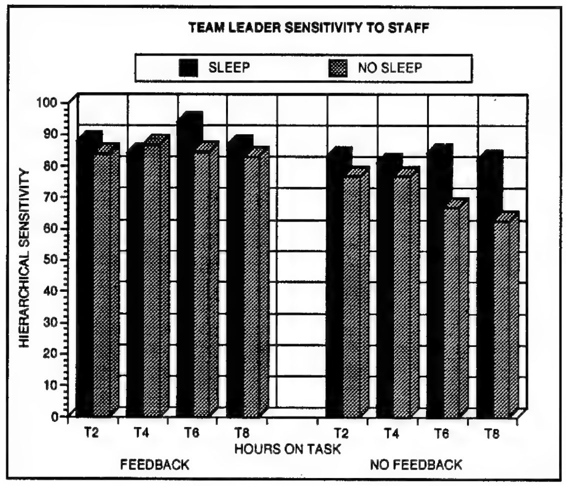


Figure 3. Team Leader Sensitivity to Staff

Similar outcomes were found at the Subordinate level of the Hierarchical Team problem. Figure 4 shows that averaged zero order correlation between subordinate judgments and true criterion scores, which captured the validity of the judgments of the criterion, became dramatically smaller over time during the No Sleep-No Feedback condition combination. Once again, the impairment was caused by a progressive deterioration in the consistency in weighting and integrating criterion information (i.e., executing the trained judgment policy), and not by a degradation in knowledge of the task (Ref 8). Here, the subordinates manifested a high level of task knowledge, however they became progressively worse at executing the task (Ref 9).

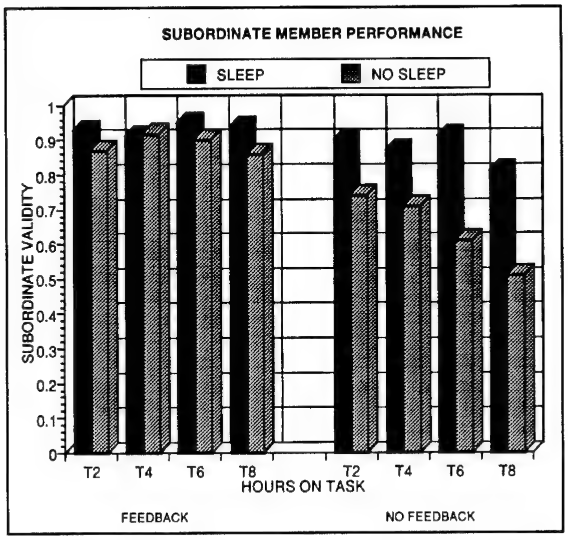


Figure 4. Subordinate Member Performance

Finally, in Figure 5 the agreement between subordinates on judgments of identical criterion values (threat values) provided some information on the progressive deterioration of subordinate judgment performance. The average zero order correlations between subordinates' judgments of the criterion were calculated. Again, the most pernicious effects on agreement were observed in the No Sleep-No Feedback condition combination. Further, on average, the agreement among subordinates became progressively worse.

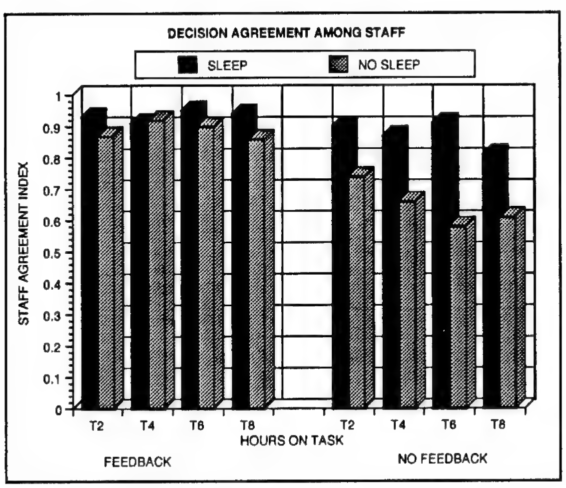


Figure 5. Decision Agreement Among Staff

5. DISCUSSION

Sleep loss, continuous performance, and delayed feedback on the outcome quality of complex decisions are all features of many operational environments. Further, these variables have been shown to consistently influence numerous aspects of performance, leading to slowed reaction time, failure to respond when necessary, false responses, slowed high level cognitive functioning, diminished working memory, and decrements in the execution of knowledge to name only a few (Ref 3; Ref 10, Ref 8; Ref 11; Ref 12). It is also evident that as little as one night without sleep can be one of the most disrupting factors in cognitive functioning (Ref 3; Ref 12; Ref 13). Thus, it is not surprising that in the present study sleep loss coupled with continuous performance and an absence of feedback was associated with a number of cognitive decrements.

There is little doubt that providing information on the quality of decision making improves performance, especially in uncertain and complex task environments. In addition, feedback reduces the "out-of-the-loop" performance problem that leads to operator failure at problem detection and control occurring when operators lose their ability to understand the complex features of an automated environment. There is even some preliminary data to suggest that feedback has the capacity to attenuate the effects of particular physiological stressors in complex judgment (Ref 14).

Further, there have been numerous studies showing the effects of withholding feedback on complex multi-information integration tasks. The interpretation frequently provided (Ref 8; Ref 15; Ref 9; Ref 16) for this decrement involves the loss in control of the execution of information. In complex judgment tasks that require a combination of analytical and intuitive

skills, resources are called upon to maintain the calibration of the organizing principle used in the integration process (i.e., the learned judgment policy). Uncertain tasks that require multiple-cue integration demand both analysis (explicit computation) and intuition (implicit computation). The uncertainty means that during training one is not completely sure of the rule (algorithm) governing successful judgments of criterion values. While some deliberation is necessary for rendering judgments (i.e. analysis), an intuitive component is also necessary (holistic assessment), and it is this component that affects the control of execution of the judgment protocol. Here, the judgment protocol has an implicit element that makes conscious awareness and explicit control over the information difficult to achieve (Ref 17). Decision makers must rely on some degree of implicit control over the information. Implicit control is enhanced by providing decision makers with immediate feedback on the quality of judgments (Ref 18).

The implications of this finding are that as team members become tired they lose the ability to control the execution of the information used in the judgment process. However, this loss is not necessarily manifested as a constant change in protocols across members. That is, individual differences in the effects of loss of control result in no two individuals being identical in the manner in which they modify their cue weighting policies over time (i.e., reduced agreement). Hammond and Grassia (Ref 19) have indicated that people often disagree about the facts, the future, the value, and action (what to do). While it is difficult to generate definitive conclusions regarding the process underlying changes in how team members perceived the decision problem over the course of the experiment, this study suggest that the process associated with weighting and integrating uncertain information maybe particularly sensitive to the effects of sleep loss and the absence of feedback. Over time, and in the absence of calibrating information (feedback) subjects dissociate from a common judgment policy they were trained to use and configure individual policies that are uniquely invalid.

6. ACKNOWLEDGEMENTS

This research was sponsored by the Air Force Office of Scientific Research. Although the support for this work is gratefully acknowledged, the ideas and comments expressed herein are those of the authors and are not necessarily endorsed by the funding agent.

Our special thanks go to D. Rocky Perez for his careful review of editing and completion of this article.

7. REFERENCES

1. Tepas, D. I. & Mahan, R. P., "The many meanings of sleep." Work & Stress, 3(1), 1989, pp. 93-102.

2. Krueger, G. P., "Sustained work, fatigue, sleep loss and performance: A review of the issues." Work and Stress, 3(2), 1989, pp. 129-141.

3. US Congress Office of Technology Assessment, "Biological rhythms: Implications for the worker", Washington, D.C., U.S. Government Printing Office, OTA-BA-463, 1991.

4. Mahan, R. P., "Extended operations and soldier judgment: A human-system design challenge." U. S. Army Aviation Digest, 1-90-1, 1990, pp.11-16.

5. Tepas, D. I. & Monk, T. H., "Work Schedules." Handbook of Human Factors. G. Solvency, New York, John Wiley & Sons, Inc., 1987, pp. 819-843.

6. Hollenbeck, J. R., Sego, D. J., Ilgen, D. R., & Major, D. A., "Team interactive decision exercise for Teams Incorporating Distributed Expertise (TIDE^2): A program and paradigm for team research." Technical Report No. 91-1, Washington, DC, U.S. Government Printing Office, 1991.

7. Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., Phillips, J., "Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise." Journal of Applied Psychology, 80(2), 1995, pp. 292-316.

8. Mahan, R. P., "Stress induced strategy shifts toward intuitive cognition: A cognitive continuum framework approach.", Human Performance, 7(2), 1994, pp. 85-118.

9. Hammond, K. R. and D. A. Summers, "Cognitive control." Psychological Review, 79, 1972, pp. 58-67.

10. Mahan, R. P., "Effects of task uncertainty and continuous performance on knowledge execution in complex decision making.", International Journal of Computer Integrated Manufacturing, 5, 1992, pp. 58-67.

11. Hammond, K. R. and J. K. Doyle, "Effects of stress on judgment and decision making", Boulder: University of Colorado, Center for Research on Judgment and Policy, 1991.

12. Dinges, D. F., Whitehouse, W. G., Orne, E. and O. C., M. T., "The benefits of a nap during prolonged work and wakefulness." Work & Stress, 2(2), 1988, pp. 139-153.

13. Wilkinson, R. T. and R. Gray, "Effects of duration of vertical vibration beyond the proposed iso "fatigue-decreased proficiency" time, on the performance of various tasks." Vibration and combined stresses in advanced systems: AGARD conference proceedings No. 145, H. E. von Gierke, North Atlantic Treaty Organization, B19-1—B19-5, 1974.

14. Searcy, C., "The effects of continuous performance and task uncertainty on risky multicue probability judgments.", MS. Thesis, Athens, GA, University of Georgia, 1994.

15. Brehmer, B., "Learning complex rules in probabilistic inference tasks." Scandinavian Journal of Psychology, 17, 1976, pp. 309-312.

16. Hammond, K. R., "Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice.", New York, Oxford University Press, 1996.

17. Balzer, W. K., M. E. Doherty, & Oconnor, R., "Effects of cognitive feedback on performance." Psychological Bulletin 106, 1989, pp. 410-433.

18. Mahan, R. P., Kirshenbaum, S., Jilg, J. M., & Marino, C. J., "Supporting diagnostic judgment during delayed feedback decision making through reliability feedforward techniques.", International Journal of Cognitive Ergonomics, 2(3), 1998.

19. Hammond, K. R. and J. Grassia, "The cognitive side of conflict: From theory to resolution of policy disputes." Applied Social Psychology Annual, S. Oskamp. Beverly Hills, CA, Sage, 6, 1985, pp. 233-254.

Establishing Common Ground and Competence in Team Performance

John Campion
Soft-System Developments Ltd
Cross Winds, Carron Lane
Midhurst, West Sussex
GU29 9LB, UK

George Brander and Eva Koritsas
Centre for Human Sciences
DERA, Portsdown, Fareham
PO17 6AD, UK

1. SUMMARY
The work reported here forms part of a three-year programme of research to establish principles of command decision support. The part we report on here relates to the executive functions of decision management within teams. A team structure is defined as the set of authorities, abilities and resources assigned to each team member, coupled with the task-dependent relationships existing between them. The concepts of "model for a purpose", "common ground" and "balance of competences" are introduced to replace the simplistic idea of "shared mental models" in explaining team behaviour. A laboratory demonstrator system (FITMASS) is described which has been built to embody these concepts.

2. INTRODUCTION
The work reported here was undertaken for the UK Defence and Evaluation Research Agency, Centre for Human Sciences. Acknowledgement is due to our colleagues, Tracy Milner and Phil Moore who assisted in later parts of the work. The work formed part of a three-year programme establishing principles of command decision support and (more latterly) task management support within command team structures. Both land and sea systems have been addressed.

We have been developing our theory of decision support and management alongside the construction and evaluation of laboratory demonstrator systems. The theory has been embodied in an evolving decision making model, named after its three distinct components – Structure, Executive and Process (STREP). It has been described more fully elsewhere [1]. It is the management aspect of the Executive that we focus on here.

3. TEAM STRUCTURE
The simplified team structure we have been working with is shown in Figure 1, which shows the main roles within a naval destroyer command team.

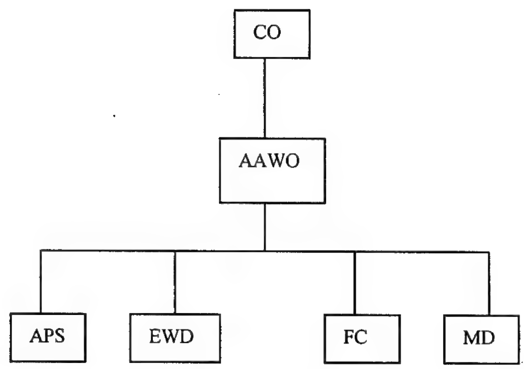


FIGURE 1 Team Structure-1

We can see here the basic three-layered structure of command. Middle command (here the AAWO or Anti-Air Warfare Officer) refers up to a higher command (here the captain or commanding officer) and refers down to a tier of directors. On the left we have the essentials of the picture compilation team (Air Picture Supervisor and Electronic Warfare Director) and on the right we have the essentials of weapon resource deployment (Fighter controller and Missile Director).

Of course, this is only a simplified broad structure. In reality there exists a complex set of delegations of responsibility and auditing which define the true and complete structure of the team. For example, the AAWO may have been given the responsibility by the CO to engage clearly hostile aircraft with his missile system subject only to his veto. The APS may have been given the responsibility to assign standard identities to tracks subject only to the AAWO's veto. In the case of other tasks, the subordinate may be required to seek approval for some action or the responsibility may be retained by higher command.

The distribution of responsibilities is fully reflected in three features of a team member. These are the *authority*

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

a team member has, the *resources* he has to carry it out and the *ability* he has to do so, as shown in Figure 2. Normally these parameters do not have fixed and discrete values, but vary across a number of team members, although there will be some cases where this does not hold. As systems become more automated and workstations more remote and general purpose, we may expect the growth towards greater distribution of competence.

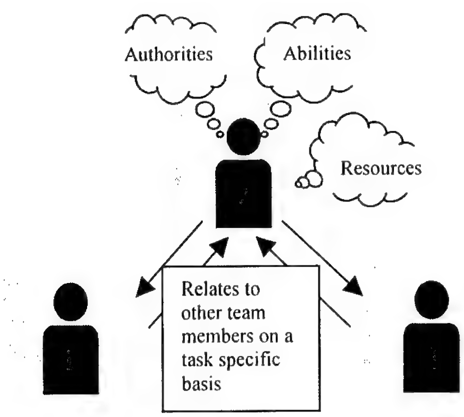


FIGURE 2 Team Structure-2

4. TEAM MANAGEMENT

The above describes the *structure* of a team, but its actual behaviour is governed by two other factors. These are, respectively, the scenario of events that it has to deal with (i.e. the nature of the work domain) and the *management* of those events within the given team structure. We shall deal first with management. There is a commonly held view that team members work well together when they have a shared mental model of the situation. However, this view requires modification and extension on three counts.

First, team members all require models of the world which are "true", but they do not have to be the "same". This is because different team members have different purposes with respect to the world and therefore require different models of it. This follows from Johnson Laird's account of models in which he describes the different models required of a TV set by a user, repairer and designer, respectively [2]. The approach introduces the concept of "truth for a purpose" which has important epistemological implications as well as practical design ones. Figure 3 shows the different views of the world held, respectively, by an EWD, MD and CO.

Second, to the extent that models are shared, the mere possession of them is insufficient in itself because effective collaboration between team members requires that collaborators have *mutually held* knowledge of the situation. That is, they not only have to have the model, they have to know that each other has, and know that each other knows and so forth. Collaborators therefore have to establish what linguists call "common ground" which is mutually held knowledge sufficient to enable effective communication.

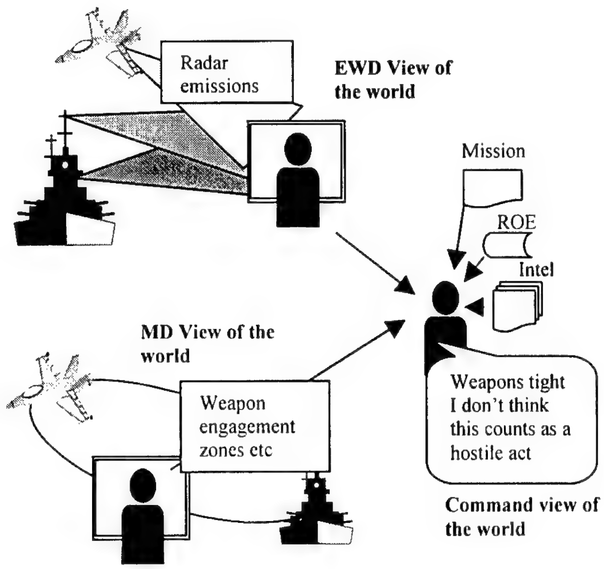


FIGURE 3 Different Views Held by Command Team Members

Third, in order to establish this "common ground" collaborators have to engage in meta-level discussions and negotiations which form part of the task management processes which many theories and models take no account of. Although we have not tried to quantify it, our observations suggest that a very large part of inter-team member communication is of this sort. How do these issues relate to our notions about team structure ?

Team members, not only are unlikely to have the same view of the world, they are likely to be differentially capable of dealing with it. The issue is not, therefore. About ensuring that everyone is equally competent but ensuring that everyone recognises the distribution of competence. Thus a command team can work well with a weak AAWO and strong Director if this situation is mutually recognised and responded to appropriately. What is happening is that the micro-structure of the team (as defined) is being altered informally to match authority to ability. This is what I call *team management.* But, as I said earlier, this has to be tailored to the nature of the work domain it is interacting with.

5. THE WORK DOMAIN

A major issue in command is the level at which control is retained. There are two philosophies – mission directed, and command directed operations. In order to establish the relative merits of each, we have not only to consider the relative competences as defined but also the nature of the work domain (the military situation) and its evolution. This is especially important, given the move towards rapid manoeuvre warfare.

Missions, especially complex ones such as joint operations, are prone to error. This is because of failures in systems, changes in circumstances and the impact of unforeseen contingencies. At present, missions are planned in great detail at the highest level. This has led to pressures to introduce a more devolved command structure (mission command) in which a unit may be given a mission and resources and left to get on with their own detailed planning. The claim is that this will lead to greater flexibility in operations and reduced failure. However this situation needs looking at in greater detail using our framework.

One of the merits of centralised command is that the greatest expertise (ability) may be concentrated in one place. However such expertise is only useful if it can be applied to good effect through having access to good information and good control mechanisms. Also the single point command must be able to handle the volume of information presented to it. If these things hold good (i.e. good communication channels and the ability to handle large volumes of information) then centralised command is good.

But we need to factor in another matter, which is the speed of response necessary. This is governed purely by the pace of external events. The aim is to be able to gather information and respond to it faster than the external events are unfolding. This is often described as getting inside the enemy's OODA (Observation, Orientation, Decision, Action) loop. This is illustrated in Figure 4. The principle is that abilities, authorities and resources should always be balanced and placed, within constraints, at such a level that the above may be maintained.

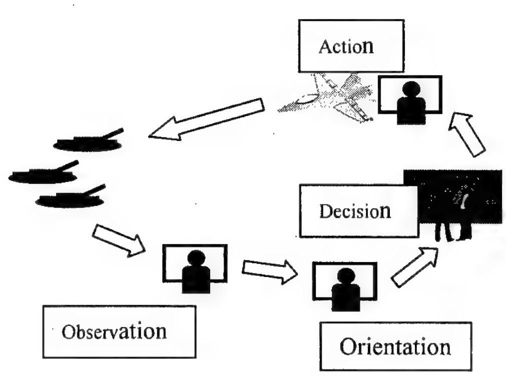


FIGURE 4 OODA Loop

Thus, the quality of the information channels can only be judged relative to the demands of the work domain. The demands of the work domain are expressed, not only in terms of the speed of their evolution, but also in terms of their "type". For example, a mission may require very particular expertise which can only be located in one place, or it may be politically highly sensitive and require that all actions are easily relateable to higher level concerns. Coupled with these, the situation may be uncertain and constantly changing. All of these factors would argue for centralised command structure. This is because a single point command *supported by very good information and control systems* would be able to respond maximally fast to a rapidly changing, politically sensitive uncertain domain.

So, in summary, we can say that a team should be organised so that there always exists a balance between the ability, resources and authority of the various units within. In addition, we can say that the locus of *management* control should be such as to relate to the nature of the work domain and the quality of the information and control systems.

6. META-COMMUNICATIONS

Taking forward our notion of common ground, it is clear that this needs to exist, not only concerning the situation, but also about the structure of the team (as defined). It is also clear that ways of establishing this common ground need to be established and supported. We have been exploring this issue at DERA.

We have developed a system called FITMASS (Fully Integrated Task Management Support System) which reflects these ideas. Figures 5 and 6 show the basic features. Space does not permit us to describe any more than the basic concept. Much of this structure reflects a management model, which is described elsewhere [3].

Each of the six team members has a workstation with the layout shown in Figure 5. In addition the AAWO has a large (full screen) PPI, main track table and the ability to alter any information on any track. As the reader can see, the support system consists of a filtered "Interest" PPI and track table which presents the basic information in the traditional way – but only for a filtered subset of tracks which are of tactical interest. At present this filtering is done manually by the AAWO.

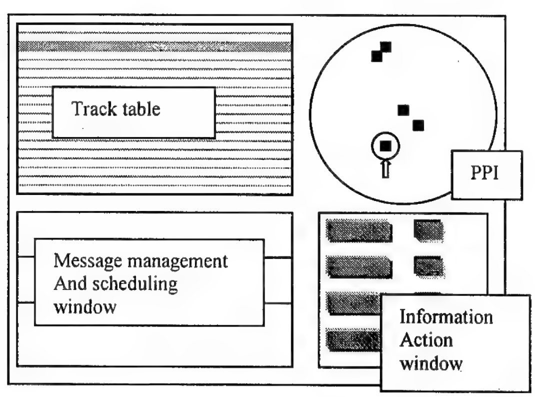


FIGURE 5 Workstation Layout

If a team member wishes to add information to a track or take action against a track he selects the track from the PPI or the track table and then indicates the information or action via the Information Action Window in the lower right quadrant.

Instead of this action taking effect directly, it sets up the appropriate messages determined by the team structure (defined in the planning phase) which seeks approval or merely inform the other team members thus identified. A standard example is if the AAWO wishes to intercept a hostile track with an outer screen aircraft. He selects the track from the PPI (or track table) and selects "Send CAP" from the IAW. This sends both a "seek approval" message to the CO and an "inform" message to the FC. The CO approves the request, which is received by the AAWO and the AAWO then orders the FC to intercept the selected track. Greater detail may be seen in Figure 6 which shows the detail of the Messaging and scheduling window.

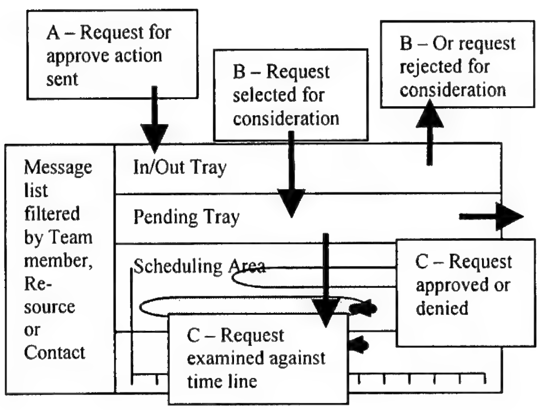


FIGURE 6 Message Management

A message is first received in the top part of the window (the in/out tray) as text in black. This can either be discarded immediately or selected for consideration, in which case it passes into the pending tray of the recipient. The text in both sender and recipient windows turns yellow. The recipient is paying attention to the message; the sender knows this and the recipient knows that he knows etc. The recipient can then, either approve the request (if that is what the message is about), reject it, or select it for entry into the scheduling window, where it may be compared with other requests. This window shows, for example, time windows for missile engagement against selected threats. From here the message may, again, be either rejected or approved. If there is some complication which needs discussion he can signal this also.

Two other important features of the system (not fully implemented yet) are that any team member can look at any other team member's message management window and can also look at the message history from a team member, resource or track point of view.

7. CONCLUSIONS

The evaluation of the system has yet to be completed. It looks at present as if the system would be useful in replacing voice communications in some cases but not in all. For example, complex threat assessments that the command wishes to pass to the whole team are not amenable to the treatment described here, but simple, stereotyped and procedural messages such as those associated with weapon assignment, are.

Much of the limitation associated with machine communication lies in its inherent inflexibility and the time and effort consumed in using it. The fast pace of complex ship air defence tends to bring out these limitations. It is likely that it would come into its own more at higher levels of command where there is more emphasis on planning and review in genuine management environments where the pace is slower and more reflective.

It also looks as if the more a team is physically distributed, the more a system such as this would be useful. It has to be said that the rich variety of ways that common ground may be established in teams who have close and direct visual and voice communication with each other, is very difficult to replace by computer means.

The system is an example of how a theoretical model can be used in a variety of ways to guide the development of a system.

8. REFERENCES

1. Campion, J., Koritsas, E., Brander, G. and Stewart, K. "A Task Model to Develop Principles of Decision Aiding", presented at ICCRTS '96, Bosworth, UK, September 1996.

2. Johnson-Laird, P. N. "Mental Models", Cambridge, Cambridge University Press, 1983.

3. Campion, J., Koritsas, E. and Brander, G. "Modelling and Supporting Interactive Behaviour and Performance in Military Teams", presented to NATO Research Study Group 4 Workshop on Test and Evaluation, Brussels, June 1997.

Shared situation awareness in Low Intensity Conflict (LIC) settings – Typical problems and emerging support solutions

R G Pascual and M C Mills
Centre for Human Sciences (CHS),
Defence Evaluation Research Agency (DERA).
DERA Fort Halstead, Sevenoaks, Kent TN14 7BP, UK

Abstract

This paper addresses the problems associated with maintaining shared situation awareness and other effective team processes in military structures operating in Low Intensity Conflict (LIC) settings. Typically, these teams are both ad-hoc in nature and distributed in nature. These issues are illustrated using evidence recently gathered from questionnaires and interviews, targeted at military personnel with experience of operating in these team settings. Conclusions drawn from this and other related DERA studies are being used to scope and direct emerging training and technology support imperatives. The results of these initiatives to reduce the negative impact of such problems and to support effective team performance are presented.

Introduction

The work grew out of a number of general observations and from findings emerging out of a number of related research programmes.

The demands of large, complex military tasks have led to the increasing use of distributed and ad-hoc team structures. This is typified by coalition warfare (as in the Gulf), and in multi-national peacekeeping settings, such as Bosnia. However, even in traditional task environments, such as warship damage control, there is a reliance on the effective functioning of many distributed sub-teams, who have to maintain shared situation awareness over multiple electronic communication networks.

There is an increased requirement for survivability, especially to maintain robust and resilient command and control structures, thus requiring the distribution of valuable human resources. In addition, there exists a requirement to gain access to expertise, advice and information that are not normally available. This is especially relevant in LIC settings, where access to legal and media specialists and linguists is essential to effective team performance.

Drawing on recently obtained questionnaire and interview data, the remainder of this report will initially describe some of the difficulties faced by teams in these settings. This will be followed by a discussion of a number of emerging support solutions to target the problems highlighted.

Typical characteristics of LIC settings

There are a number of readily identifiable characteristics of LIC settings that combine to make effective team interaction difficult.

Unlike more traditional cohesive units, the teams operating in these environments are often composite forces. At the macro level, they are composed of units from many different countries, bring cultural and linguistic barriers to effective communication and co-ordination. At the micro-level, these teams now have added specialist personnel, such as legal and media specialists, who are often only available for short periods of time. The overall result is that team norms and identity are slow to evolve, requiring more effort to achieve effective team co-ordination and cohesion. This is exacerbated by the lack of training opportunities for these ad-hoc teams (frequently less than six months), which also serve to disrupt the formation of accurate shared teamwork mental models.

In LIC settings, the neutral role adopted by forces often means that their numbers are deliberately not overwhelming, and thus they are often geographically distributed when performing their operational duties, e.g. patrolling borders. This leads to remoteness within the team and makes it more difficult to maintain shared situation awareness and to implement important teamwork behaviours effectively.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

The rapid dynamics underpinning the evolution of many modern conflicts can result in a lack of adequate contingency plans. The subsequent implementation of hot plans tends to produce confused, fluid mission statements. This poses additional challenges for team leaders when specifying team goals, potentially introducing ambiguity into teamwork planning.

Decision making and effective team process is also made more complex because of conflicting goals, different interpretations of situation information and rules of engagement and cultural differences in *modus operandi*.

In these environments, the traditional Strategic-Operational-tactical (SOT) team hierarchy has been replaced by flatter structures, where the influence of the team leader may be diminished, and individuals on the ground have to take more decision making responsibility, with potential strategic consequences.

In summary, the combined impact of the factors described above is that teams are likely to find it difficult to maintain a shared level of understanding of the situation dynamics and a shared vision of team goals. In addition, teams also have difficulty in developing and maintaining appropriate shared mental models. It is suggested that the degree to which a team will maintain high levels of shared situation awareness is strongly related to its ability to evolve and support accurate shared mental models. These are knowledge structures collectively held by the team that enable them to co-ordinate their actions and to anticipate future member requirements. Strongly shared mental models promote synergy between effective taskwork and teamwork, and the importance of this concept both to team problems identified and emerging solutions will become clear.

Survey and interview analysis

As part of the DERA research programme, a questionnaire was developed and targeted at military personnel who had operated in ad-hoc and distributed team structures, to establish their views on all aspects of teamwork in these settings. A sample of these personnel was subsequently interviewed in detailed knowledge elicitation sessions.

The survey is ongoing, and to date the responses from 32 individuals have been analysed. The respondents represent all three services, with the majority being land-based personnel. The questionnaire responses also reflect a wide range of experience, from Brigadier through to Lance Corporal and Able Seaman. More than 80% of respondents had experience of LIC settings, with 60% having operated in teams that were both ad-hoc and distributed.

Table 1 illustrates the mean responses for a sample of the critical issues addressed in the survey.

Key problem areas	Identified by	
	Ad-hoc team members	Distributed team members
Communications	97%	66%
Establishing & maintaining SSA	83%	72%
Additional leadership challenges	62%	86%
Engaging in core teamwork behaviours	75%	71%
Lack of training	79%	79%

Table 1. Key problem areas identified

As can be seen, establishing and maintaining situation awareness, leading teams effectively and engaging in essential teamwork behaviours (such as monitoring and offering feedback) were all described as particularly difficult by a substantial majority of respondents. Significantly, only 11% of respondents considered that existing training regimes were totally sufficient for preparing teams to operate most effectively in both ad-hoc and distributed settings.

Two of the key problem areas identified by survey respondents are discussed in more detail below, beginning with a review of the challenges faced by team leaders in LIC settings.

Team leadership

The nature of the problems experienced by team leaders can perhaps best be summed up by this comment drawn from an interview with a British Army officer;-

'In a situation I faced in Kuwait, I was in charge of more than 300 people from many nations who were literally thrown together. I didn't know any of them, none of us really knew what we were going to do. I didn't know what level of training or particular talents any of them had, I didn't know what made them tick, so when I came to distribute them, I had no idea who would make a good partner with who, who would survive in more rugged areas, and who wouldn't…'

The team leader has a critical role in ensuring co-ordinated team and task performance and to be

effective they must foster respect with the team, act as a role model, and to provide and receive feedback within the team.

As one respondent to the DERA survey noted, 'the greatest challenge to any leader is creating a team from a group of individuals'. Leaders of teams must provide a model of teamwork, which promotes leader perceptions of goals, roles and responsibilities, and expectancies concerning future team responses. This model can be difficult to communicate to ad-hoc team members who may bring different perceptions of teamwork, situational understanding, goals and in multi-national teams, even conflicting political agendas.

In ad-hoc and distributed settings, it is more difficult for team leaders to exercise their personal style and to foster respect among remote, or relatively unknown, team members'. There exist special challenges in knowing distributed team members, their areas of expertise and how to best exploit them for effective team performance.

It is also more difficult to monitor performance and to offer feedback in distributed settings. Available communications media tend to result in reduced information richness. Although the team leader has sight of team products, they may not necessarily see the process by which these products were arrived at. Team leaders in LIC settings must therefore pay additional attention to explicitly implementing mechanisms for the distribution and reception of situation assessment updates.

Survey respondents also argued, that when operating in LIC settings, team leaders need to be particularly flexible and have heightened interpersonal skills. In addition, the team leader must have the ability to rapidly integrate new team members into what are often dynamic changing group structures.

Implementing core teamwork behaviours

The other main area addressed by respondents were the difficulties surrounding the implementation of core teamwork behaviours, such as the monitoring of colleagues or the provisions of feedback within the team [See 1 for a full review of key teamwork principles].

In order to maximise team efficiency and to build up a psychological contract of trust among team members, the monitoring of fellow members' performance is considered critical [1].

Monitoring in distributed teams has to be largely conducted using various electronic links. Unfortunately this results in a reduction of cues to act upon, making accurate situation assessment of team and system states more difficult [2]. Many team leaders commented in the survey that it was easier for people to hide their anxieties and problems.

With ad-hoc teams, it was argued that individuals might be reluctant to monitor, or may be unaware of the need to monitor, their colleagues. Ad-hoc teams have the additional problem that they typically have limited opportunities to train together and thus build up the shared mental models that contribute to the establishment of team identity. Consequently, knowledge of fellow members' strengths and weaknesses is limited, and team interdependence is likely to be weakened. The provision of accurate and timely information feedback within the team is critical for building up adequate levels of shared situation awareness.

As with the process of monitoring, survey respondents noted that an ad-hoc team with an immature sense of team identity is likely to find it more difficult to provide and accept feedback. Unfamiliarity with fellow team members in ad-hoc teams also contributes to the difficulty of interpreting the meaning of restricted response information. It is suggested that this may, in part, be a function of inconsistent mental models held by team members, leading to mistaken assumptions concerning recipient understanding of feedback data.

Distributed teams in LIC settings also have similar problems, because of the reduced quantity and quality of cues available across electronic media, from which to try and develop appropriate responses. Reduced feedback within the team undoubtedly impacts on their ability to maintain a high level of shared situation awareness.

Survey respondents were also asked about the degree to which teams engage in backing-up behaviour during LIC operations. Many argued that it was more difficult for members of ad-hoc teams to know how and when to support (i.e. provide assistance to) each other, compared with mature teams. One officer commented; 'there are natural inhibitions between people who are unfamiliar with each other. This is exacerbated where the ad-hoc team is joint, combined, or coalition in nature...where the more subtle uses of language are unfamiliar to one party and may be lost to other team members'. Ad-hoc teams do not have well-developed norms, or levels of trust, that facilitate the practice of backing-up

behaviours. Individuals do not know the strengths and weaknesses of their colleagues and, as was also observed, 'are naturally preoccupied with orienting themselves to the new task/team environment'.

Similar problems were reported for distributed team structures, though for different reasons. It was argued that even if team members have an increased sense of team maturity and interdependence, restricted modes of communication degrade important cues concerning stress, fatigue and workload. Thus, difficulties can arise because of a lack of pertinent information upon which to base accurate judgements, thus delaying timely interventions.

Emerging support solutions - Teambuilding

Collectively, the findings from the DERA survey and follow-up interviews, highlighted that team members in LIC settings have diminished shared mental models for team goals, roles and interdependencies. It was also suggested that these teams tend to perceive themselves as 'loose collectives' as opposed to integrated, directed and cohesive units.

Research conducted with the University of Surrey has been examining the value and potential applicability of team-building techniques for enhancing the effectiveness of teamwork in LIC settings. To this end, 40 UK team-training consultants were recently interviewed to obtain an in-depth view of available methods and perspectives within this area [3].

A full review of the outputs of this work is outside of the scope of this paper. However, one central conclusion was that most existing teambuilding models examined were found to focus almost exclusively on promoting effective team behaviours. There were few references to any cognitive dimensions underpinning the interventions described.

As a result, a conceptual framework was developed, incorporating cognitive and motivational dimensions of team effectiveness when defining team-building requirements [3].

In summary, the approach adopted proposes that teambuilding should aim to promote team focus and orientation. In parallel, the teambuilding process should progress team competence at the cognitive (shared mental models), meta-cognitive (higher order self-regulatory strategies) and motivational (team identification and team potency) levels.

In building team focus and interdependence, teams should be goal directed at the individual and the team level. This orientation will be enhanced through the promotion of knowledge of team function and contribution to wider organisational goals. It is also suggested that teams should have knowledge of critical team principles, and be able to reflect on what will promote optimal team functioning. The development of this knowledge, predicated on dynamic feedback from specific targeted exercises, should be initiated at an early point in military training and reinforced throughout career development.

It is also imperative that perceptions of co-operative interdependence are established within the team, either prior to deployment, or through directed exercises 'in-situ'. This effort should anchor interdependence in the way the task is organised and the behavioural imperatives that it produces. It is also suggested that interdependence can be emphasised through the creation of clear super-ordinate goals and in the way that team success is defined.

Providing a forum in which the team can develop an explicit and realistic knowledge of itself as a team can enhance the process of building team competence. Not only does this include knowledge of team imperatives, but the sharing of information on roles, goals, operating culture, skills and team strengths and weaknesses. It is suggested that this process will be expedited through the analysis of shared mental model states within the team. A potential method for achieving this will be touched on shortly.

A forum must also be provided in which the team can develop self-regulatory skills. This means that the team can diagnose the process requirements of different problem situations, and analyse and scope out problems of both a task and team nature. A team should also aim to monitor and manipulate its own processes in a self-reflective way, so that knowledge is not only captured, but also maintained in the light of ongoing team turnover. A successful teambuilding environment should focus on active team self-review, to further promote team interdependence, feedback and the identification of behaviours resulting in team effectiveness. This process should also enhance the effectiveness of shared mental model utility, through improved knowledge of team member expectations.

Emerging support solutions – Leadership

A number of support imperatives have been generated in examining the potential support requirements of team leaders in LIC environments.

It is important to raise awareness of the behavioural cues that a team leader should monitor that would, for example, enable a team leader to know when the team is performing sub-optimally. It is suggested that these could be incorporated in summary commanders guides, or 'teamwork aide memoires'. The format of such guides could be similar to military pocket guides that are currently available to support taskwork, as in guides to producing the command estimate. Teamwork guides could cover different topics, such as teambuilding, or team self-review, as in the mocked-up example shown in figure 1 below.

Teamwork Aide Memoire
'Team Self-Review'

- What KSA's did each member bring ?
- What role was adopted by team leader & how did this change ?
- Did leader provide mental model for team ?
- Did leader monitor & co-ordinate activities of team ?
- Who briefed who and how ?
- What info was missing, conflicting or ambiguous ? Was this made explicit ?
- Who checked on constraints against plans ?
- Were roles & responsibilities defined ?
- What happened in margins ?
- How was initial team SA built up ?
- What assumptions were made ? How were these monitored/tested/updated ?
- How was planning process conducted ?
- What assumptions were made ? How were these monitored/tested/updated ?

Continued overleaf

Figure 1. Mocked-up teamwork aide memoire

As has been suggested in previous research [4], it is important for team leaders to implement explicit mechanisms for ensuring that information, particularly situation updates, are made available on a timely basis up and down the team structure. Team leaders play a critical role in facilitating shared expectations for the mission, goals, tasks, and teamwork, in order that the team can rapidly develop accurate shared mental models. In addition, the team leader is the catalyst for the creation of a forum where the team can develop self-regulatory, or metacognitive skills, encouraging the review and improvement of teamwork.

Several military personnel interviewed, proposed that a 'buddy' system should be created with incoming team leaders, so that they can develop an early appreciation of the true nature of the operating environment and the likely demands and pressures associated with creating an effective team in particular LIC settings.

Finally, it is suggested that team leaders should exploit 'down-time' to conduct diagnostic exercises to expose existing levels of shared understanding within the team for taskwork and teamwork.

Figure 2 shows a mental model quadrant graph derived from a recent study [5] with Police Armed Response Units. In this example, the graph represents their collective views of what makes effective teamwork for these type of teams. As such, mental model quadrant graphs provide a mechanism for visualising the team shared experiential mental model.

The diagram is divided into four quadrants, with each numbered dot illustrating the teams' views concerning the importance (or criticality) and the amount of agreement (or consensus) for particular teamwork characteristics. The more characteristics in the top right quadrant, the greater the teams' shared mental model for the area under examination. It can be used to quickly highlight shared perceptions and potential differences in thinking between team members. Characteristics in the bottom right quadrant are those that are seen as important by some team members, but not by others. For example, if a number of team members consider it important to offer feedback to their colleagues, whilst others do not, this is an issue that can be addressed by the team leader or team trainer in developing team effectiveness. As such, it is therefore also a method for comparing leader and team perceptions and for supporting the development of training interventions.

Emerging support solutions – Technology

It is also suggested that further research needs to be conducted to examine the ways in which Computer-Supported Co-operative-Working (CSCW) technology can be used to support teams shared mental model utility.

In a related DERA programme, effort has recently been devoted towards developing a research vehicle designed to support a shared knowledge environment. This has utilised a mixture of bespoke and 'off-the-shelf' software.

In using the system, each team member is encouraged to provide information in relation to core mental model constructs, such as knowledge and interpretation of individual and team goals, expectancies and resources. The team leader has a

composite view of this information and can check the consistency of team member perceptions and intentions, and initiate correcting action where necessary. The team leader can also enhance shared situation understanding within the team, by seeking additional information from a team member in order clarify meaning and to reduce ambiguity.

The system can therefore be used to improve metacognition within the team and to trigger important teamwork behaviours. For example, provoking team members to check their assumptions and goals with their colleagues, in light of inconsistencies or knowledge gaps highlighted by the team leader. Alternatively, the team leader could redefine some aspect of taskwork, or teamwork, and quickly promulgate the information to other team members.
Thus far, the system has only been trialed in a small research team setting on the DERA corporate network. However, it has generated real insight into team working, and has highlighted early on potential teamwork problems and shared mental model inconsistencies.

The research is now being progressed to investigate how mental model constructs can best be represented within the system and how such a system can be embedded within a normal task regime. It is also planned to extend the testing of a refined version of the system to a specific real-world team context.

Conclusions

In summary, the increasing utilisation of ad-hoc and distributed military team structures, typically found operating in LIC settings, poses significant challenges for effective team performance. In particular, teams in these settings are likely to encounter difficulties in evolving and maintaining accurate and timely shared mental models for taskwork and teamwork.

A DERA survey conducted with individuals who had operated in ad hoc and distributed teams, confirmed many of these assertions. Respondents noted that communications, establishing and maintaining situation awareness, implementing core teamwork behaviours (such as monitoring), and leading teams effectively, were all challenging problem areas.

It was also felt that existing training courses could be enhanced to better support the needs of leaders and team members operating in LIC environments. It is suggested that teamwork training should be introduced early and systematically reinforced in existing training regimes. This should encourage the application of metacognitive skills and emphasise the importance of team interdependence in these settings. In addition, team-building techniques could also be applied once a team has deployed, to facilitate the rapid formation of shared mental models.

Team leaders should be trained to monitor specific behavioural cues, enabling the early identification of poor team performance to be recognised. This type of information could be incorporated into teamwork guides that could also cover other topics, such as team self–review. Team leaders should also seek to obtain an early understanding of mental model states within the team. A technique for obtaining this type of information has been described in this paper, that enables the visualisation of shared perceptions and inconsistencies in thinking between team members. This information can then provide an input to the development of appropriate training interventions.

Finally, it is suggested that the potential role of CSCW technology in the support of ad-hoc and distributed teams should be explored further. Initial research conducted within DERA, suggests that networked shared information environments can highlight teamwork problems and support the development of shared mental models.

Acknowledgements

The authors would like to thank the Social Psychology European Research Institute, at the University of Surrey, for the research examining the application of team-building techniques.

References

1. McIntyre, I. M. and Salas, E., (1995). Measuring and managing for team performance: Lessons from complex environments. In Guzzo, R. A. and Salas, E. (Eds.), *Team effectiveness and decision making in organisations.* Jossey Bass, San Fransisco.

2. Sproull, L. S. and Keisler, S., (1991). Connections: New ways of working in the networked organisation. MIT Press.

3. Millward, L.J. and Purvis, R.G., (1998). Team building techniques – A critical evaluation. Social Psychology European Research Institute, at the University of Surrey. Internal report prepared for DERA (CHS). [unpublished]

4. Serfaty, D., Entin, E.E., and Volpe, C., (1993). Adaptation to stress in team decision making and co-ordination. In *Proceedings of the 37th Annual Human*

factors and ergonomics Society Annual meeting. Seattle, WA: The Human Factors Society.

5. Pascual, R.G., Henderson, S.M., and Mills, M.C., (1998). Understanding and supporting team cognition – Year 1 report and recommendations. [Unpublished].

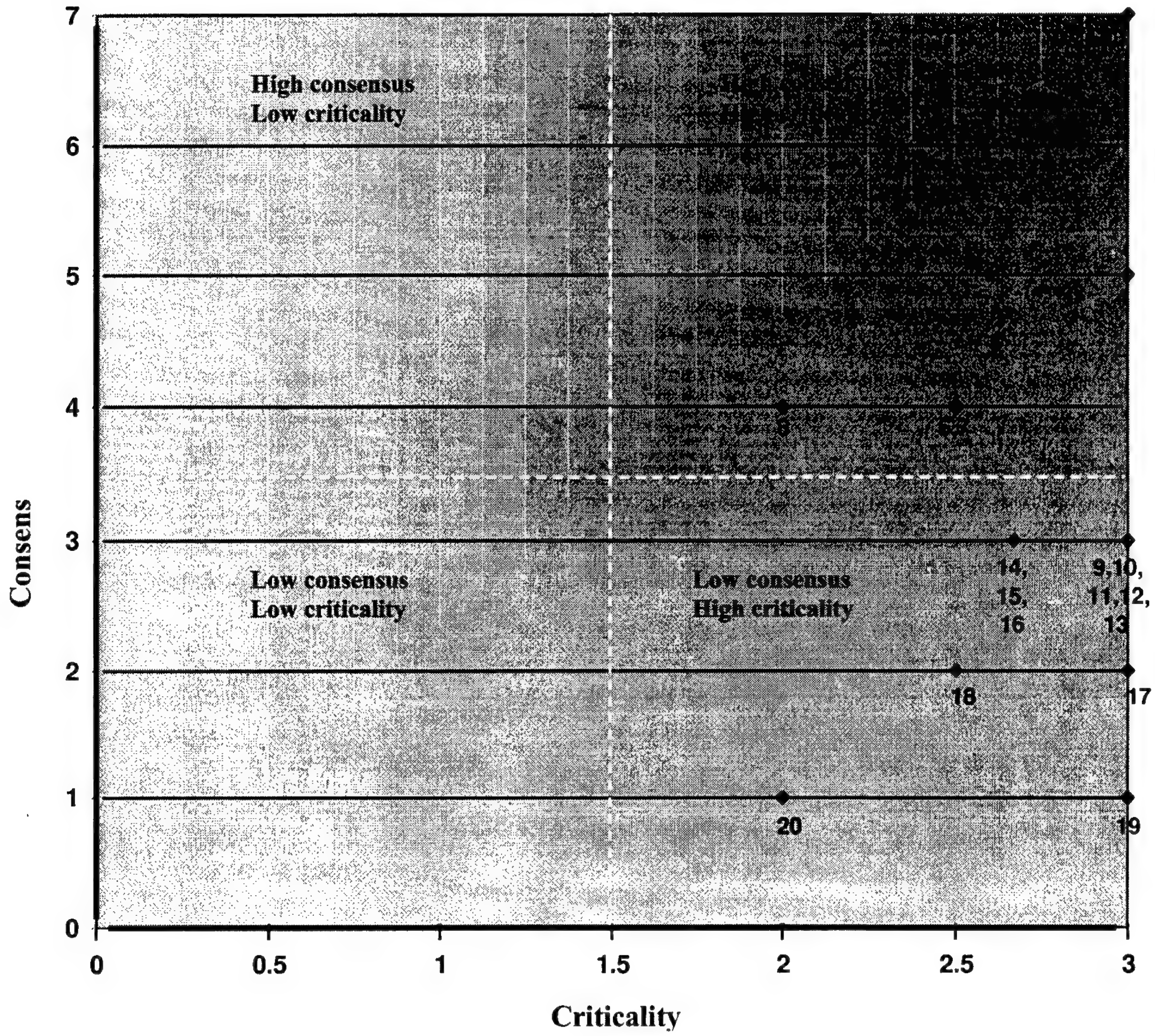


Figure 2. Mental model quadrant graph

The effects of two types of information exchange on team self-correction

P.C. Rasker, J.M.C. Schraagen, W.M. Post, & E.R. Koster
TNO Human Factors Research Institute
P.O. Box 23
3769 ZG Soesterberg
The Netherlands
Rasker@tm.tno.nl

ABSTRACT

In this study, we investigated the role of information exchange in Command & Control teams. For the sake of efficiency, it is often thought that information exchange in teams should be restricted to "what is needed". Team members are expected to exchange relevant data only. We hypothesized, however, that additional information exchange in teams would contribute to the performance. Furthermore, we expected that teams use the time during breaks between task execution to evaluate their task performance, which may improve their performance for the next time. This phenomenon is called "team self-correction". Two types of information exchange that play a role in Command & Control teams were distinguished. First, during task execution teams engage in "activity-based" information exchange: information exchange with concrete content concerning the ongoing flow of events. Second, between task execution (or in low workload periods) teams engage in "task-related" information exchange: information exchange with a conceptual and abstract content concerning the task performance in general. In an experiment, we investigated the effects of these two types of information exchange on team self-correction and the overall team performance. Four conclusions can be drawn from the experimental results. First, in order to improve team performance team members need to exchange additional information, besides the necessary data. Second, team performance improves, when team members engage in "task-related" information exchange. This strongly supports the notion of team self-correction. Third, so called "activity-based" information exchange during the task contributes more to the team performance than so called "task-related" communication outside the task. This may be explained by better opportunities for coordination, more possibilities to cross check and to correct errors, a better shared understanding of the situation, and more possibilities to learn. Finally, "activity-based" based information exchange is required to deal successfully with novel situations.

KEYWORDS

Team, Command & Control, information exchange, decision-making, communication, coordination

1 INTRODUCTION

Everywhere where people need to work together teams are found. In working situations, one could think of examples such as military teams, fire-fighting teams, airline cockpit crews, surgical teams, nuclear power factory teams, and management teams. These teams have in common that they have to work under complex and dynamic circumstances, which can be characterized by time pressure, heavy workload, ambiguous information presentation and a constantly changing environment. Moreover, teams in these settings have to face high stakes where poor performance will often have considerable consequences.

Before giving an outline of this paper, we first explain what we mean when we speak of a team. A team is defined as a set of at least two people that work together towards a common goal, who have each been assigned to specific roles or functions to perform, and where completion of the goal requires dependency among the group members (Dyer, 1984; Salas, Dickinson, Converse & Tannenbaum, 1992). According to Orasanu & Salas (1993), the dependency among the members consists of information, knowledge, and means for reaching their common goal.

1.1 Generic Command & Control functions

This paper focuses on teams that perform in Command & Control situations. Figure 1 gives a descriptive model of generic Command & Control functions that can be executed by teams. The model is based on Adams (1995) and Passenier & Van Delft (1995). It is a model that describes a closed-loop, real-time work process.

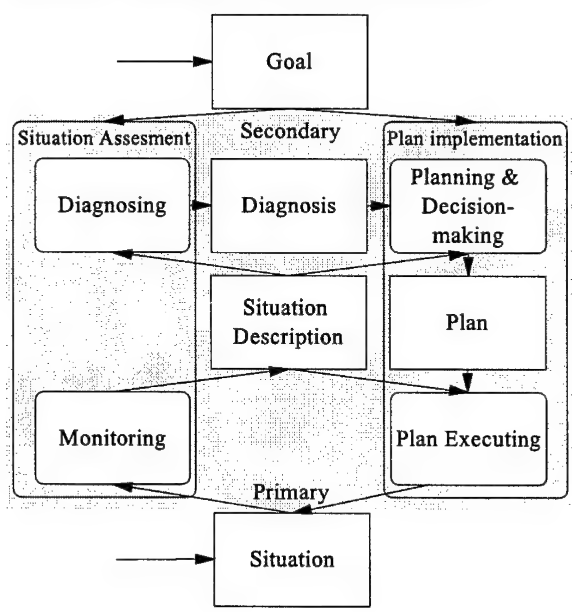


Figure 1 Generic Command & Control functions

The generic decomposition of Command & Control functions consists of a set of individual sub-functions (rounded boxes) and information units (angled boxes). The sub-functions are part of the functions "Situation Assessment" and "Plan Implementation" and are distributed over two levels of information transfer. The

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

primary level represents a direct response to a monitored event, which is comparable to the process of rule based behavior (Rasmussen, 1983). When monitored events are investigated in the light of the goal and plans are developed, then we speak of the secondary level of information transfer. This is comparable to the process of knowledge based behavior (Rasmussen, 1983).

The current situation is the input for the Command & Control process. Monitoring consists of assembling and maintaining a picture of the actual situation, which results in a description of the situation. At the primary level, a monitored event in the situation may trigger a direct response that leads to the execution of a pre-defined plan (Plan Execution). At the secondary level, Diagnosing the situation takes place in the light of the goals to be achieved. This diagnosis is used as input for the development of plans. Planning & Decision-making encompasses the initiation of activities in order to achieve the desired goal. At the secondary level higher-order objectives, determined by the goal, and the type of activities, are translated by the Planning & Decision-making function into plans for the primary level. At the primary level, Plan Execution takes account of the execution and control of activities. Changes in the situation are noticed by the Monitoring function. The situation, however, can also change by external events.

1.2 Information exchange in teams

Teams can execute the generic Command & Control functions described in the former section. Field studies have shown that good and poor teams can be distinguished based on their information exchange. Orasanu & Salas (1993) found that effective cockpit crews engage in highly task directed communication that involves plans, strategies, intentions, possibilities, explanations, warnings, and predictions. Information about intentions, task-related needs, positions, roles, needs, responsibilities, and expectations can be planned in advance. For successful teams, this takes place particularly during periods of low workload (Orasanu & Salas; Stout & Salas, 1993). Seifert & Hutchins (1992) point at three important functions of communication: information exchange, error detection and the acquisition and maintenance of a shared view (model) of the situation.

The ability of team members to give, seek, and receive task-related feedback is known as performance monitoring (see for example, Cannon-Bowers, Tannenbaum, Salas & Volpe, 1995). This includes the ability to accurately monitor the performance of other team members, provide constructive feedback regarding errors, and offer advice for improving performance. Effective teamwork requires that team members keep track of their fellow team members' performance, while carrying out their own task (McIntyre & Salas, 1995).

Team self-correction after task execution (i.e., team members giving each other feedback) supports the development of shared expectations and shared explanations (Blickensderfer, Cannon-Bowers & Salas, 1994). Team self-correction is viewed as a process in which team members engage in reviewing events, correcting errors, discussing strategies, and planning for the next time. By doing this, team members correct their team attitudes, behaviors, and cognitions. It is hypothesized that self-correction discussions after a performance session help to clarify expectations of the team and the task, which increases task understanding and fosters shared knowledge.

The ACT-R theory of cognitive functioning developed by Anderson (1993) provides us with an alternative explanation of how intra-team feedback may facilitate team performance. According to this theory, people learn cognitive skills through the acquisition of production rules, which are "if-then" or "condition-action" pairs. The "if", or "condition", part specifies the situation under which the rule applies. The "then", or "action", part of the rule specifies what to do in that situation. When people learn, a new production rule has to be created, which means that the feedback must be linked to that specific situation as much as possible. Anderson (1993) asserts that the delay between a production's application and the moment of feedback affects the rate of learning. Based on this theory it is expected that intra-team feedback gives more opportunities to improve team performance when feedback is provided during task execution rather than between task execution. The rationale behind this is, that intra-team feedback during task execution is more specific and better linked to the situation compared with intra-team feedback that takes place between performance sessions. Consequently, the learning rate will be higher when feedback is provided during task-execution rather than between task execution.

Besides the advantages of information exchange outlined above, there also disadvantages. Communication is error prone and can disrupt the workflow during high workload periods (Hutchins, 1992). For this reason, officers in naval defense feel that the amount of information exchange in Command & Control rooms is too high and should be restricted to "what is needed".

1.3 Information exchange in Command & Control teams

What type of information exchange takes place when teams have to execute the Command & Control functions (see Figure 1)? First, in order to execute the plan, all relevant data must be obtained and exchanged within the team. Moreover, functions cannot be executed without knowledge about the domain. It is expected that team members feel the need to exchange information concerning their knowledge about the domain. The final type of information exchange with regard to all functions of the Command & Control process is called meta-communication, that is concerned with the management of the information exchange between team members.

The exchange of information between team members concerning their current activities, enables the monitoring of each other's performance. Therefore, it is expected that for the Command & Control function of plan execution teams exchange information concerning their tasks. For the execution of the diagnose function it is expected that teams have to exchange information where they evaluate the result of their activities. One can distinguish two types: information exchange that is concerned with a) the evaluation of actions that are performed at that moment; and b) with the evaluation of the task in general. For the execution of the planning & decision-making function it is expected that teams have to exchange information where they plan their activities in order to achieve the goal. Again, one can distinguish two types of information exchange that are concerned with a) the coordination of activities that are performed

at that moment; and b) the formulation of general plans and strategies.

On the basis of the above analysis, nine categories of information excange are defined (see section 2.5).

Two types of information exchange may be abstracted from these nine categories. First, during task execution teams engage in "activity-based" information exchange; team members exchange information about the activities they are carrying out at that moment, give each other direct feedback, communicate their intentions, and plan future activities. This is a direct form of communication with a concrete content concerning the ongoing flow of events. Second, after task execution (or in low workload periods) teams engage in "task-related" information exchange. That is, team members exchange information about the task in general, evaluate the whole task performance, and talk about general strategies. This "task-related" information exchange can be characterized as a general form of communication with a conceptual content concerning the task as a whole.

1.4 Research question

This study is an attempt to investigate the relative contribution of the two types of information exchange as described above. The first hypothesis concerns the efficiency of information exchange in teams. Although it seems efficient to restrict the information exchange to "what is needed", we think that this decreases team performance. Additional information exchange gives teams the opportunity to cross check errors, develop and maintain a shared view, and monitor each other's performance and provide feedback, which contributes to the team performance. Therefore, it is expected that in Command & Control teams, information cannot be restricted to data only. To test this hypothesis, we compared the performance of teams that could exchange information with teams that could exchange the necessary data only.

It is assumed that teams that exchange information during task execution will engage mostly in "activity-based" communication, whereas teams that exchange information between task execution will engage mostly in "task-related" communication. No assumptions are made as to which will be the most beneficial for team performance. According to the team self-correction concept, it is expected that performance discussions between task execution will contribute to the team performance (Blickensderfer et al., 1994). Based on the ACT-R theory, however, one would expect that information exchange during task execution is more beneficial when compared to information exchange between task execution. Therefore, one undirected hypothesis will be tested: there is a difference in performance amongst teams depending on the period during which they can exchange information unrestrictedly (i.e., during versus between task execution).

2 METHOD

2.1 Subjects

The data for this experiment was obtained from 88 students of Utrecht University in 44 teams of 2 subjects. The distribution of subjects over the different conditions with regard to sex was as follows: 3 female, 3 male teams and 5 mixed teams in the "restricted communication" condition; 5 female and 6 male teams in the "unrestricted communication" condition; 6 female teams and 5 male teams in the "during scenarios communication" condition; 5 female teams and 6 male teams in the "between scenarios communication" condition. The conditions are described in section 2.4. The subjects were paid Dfl. 60, = and were informed that they had a chance of receiving a bonus of Dfl. 40, =.

2.2 Task

To study team behavior in Command & Control situations, an experimental low fidelity simulator has been developed (Schraagen, 1995). This simulator is an interactive computer game, in the form of a fire-fighting task that has to be played by a team of two operators. The task is modeled on basic characteristics of Command & Control functions (Van Delft & Schuffel, 1995). A detailed analysis of the fire-fighting task can be found in Post et al. (1997).

2.3 Outline of the fire-fighting task

In an imaginary city, an arsonist sets buildings on fire. The fire-fighting team, comprised of an "allocator" and an "observer", has to locate the fires and extinguish them in order to minimize the number of casualties, which is the overall goal of the task. To accomplish this goal several tasks have to be carried out in parallel, requiring the observer and the allocator to work together. The observer surveys the city and reports the status of buildings in the city to the allocator by sending pre-formatted messages. The allocator receives these messages and acts upon them by allocating a number of resources (i.e., fire-fighting units) to the buildings in order to extinguish the fires. Different types of buildings in the city need different numbers of units to fight the fire and are associated with different numbers of (potential) casualties.

The system with which the allocator and the observer interacted consisted of two linked workstations. The allocator and the observer each had their own computer screen at their disposal on which a graphical interface was represented. By pointing and clicking with a mouse, the allocator and observer interacted with the system.
On the interface of the observer, a schematic map of a city was represented. The map contained objects, buildings of five different types: houses, apartment buildings, schools, factories, and hospitals. The map was divided in four sectors. By pointing and clicking on buildings the observer could gather information on a particular building. This information was then displayed on a "current building message window", and could be sent to the allocator, by highlighting and clicking on "buttons". The display also contained a message overview window. Fires were indicated by a flashing red contour around the building. A green contour indicated a fire was extinguished and a thick black contour, with crossed black lines, indicated a building was burnt down.

The allocators did not have a map of the city. Their display contained a message window in which the messages sent by the observer were displayed. The allocator could forward these messages to an overview window, which allowed the allocator to manipulate the number of resources allocated to the various buildings. The allocator display also contained a "fire-station window" in which the number of fire-fighting units

available was listed.

The team played several scenarios in which different buildings were set on fire. These scenarios were equal in length (three minutes real time) and were divided in 12 periods of 15 seconds each. In each period, the status of buildings could change from no fire to fire, from fire to saved or burnt down. In addition, the number of fire-fighting units needed during the fire could change. Once a fire started, it took several periods before the fire could be extinguished, depending on the number of fire-fighting units present and the time they arrived (and stayed) at a building. Events within scenarios, that is, what building is set on fire in what period were pre-programmed. Once a fire was started, pre-programmed algorithms (so called state transition diagrams), determined how the fire developed in reaction to the deployment of resources by the team.

The allocation of fire-fighting units took some time. Units allocated from the fire station to a building needed one period to reach their destination. Since fire-fighting units always had to come back to the fire station before they could be allocated to another building, it took longer to redirect resources from one building to another than to direct them to that building from the fire-station. The allocation commands of the allocator (adding or pulling back resources from buildings by manipulating "+" or "-" buttons) were effectuated at the change of periods.

The central theme in the fire-fighting task was to detect a major building ("target building") that was most likely to be set on fire later in the scenario. The number of potential casualties involved in these buildings outweighs any other number of casualties. It was essential then, for the team to locate this building in time and to (re)allocate fire-fighting units to save people. In order to accomplish this task, team members needed to recognize patterns and find a warning message indicating which building was going to be set on fire in what period. A pattern could be abstracted from the particular sequence of three buildings catching fire earlier in the scenario. The sectors, in which these buildings were located, indicated in which sector the target building would be located, hence limiting the possible targets. The sequence and type of buildings involved in this pattern indicated the type of target building, narrowing the possible targets further down. In the period in which the last building of the pattern started to burn, the observer needed to check the four possible target buildings by clicking on the building icons on the map. The observer needed to check all possible buildings because the target building could not be discriminated from other buildings of the same type by looking at the map. When the observer clicked on the right building a message appeared in the message box, indicating "danger", the period in which the building would catch fire, and the number of units needed. This message was then sent to the allocator who could start then to re-allocate fire-fighting units. Subjects were instructed and trained in detecting and interpreting these patterns.
The prediction of the location of a target building was based on a simple rule: the sector diagonally opposite to the one in which the pattern buildings were located. The teams were instructed, however, that pattern recognition would not give a hundred percent guarantee of finding the target building in the opposite sector.

To simulate communication problems (e.g., distorted radio connections), in some scenarios the pre-formatted e-mail messages were distorted (only during the experimental task). This was done by adding a different identification label to the building. The distortion of the pre-formatted e-mail messages was fixed as follows. First, only the messages transferred from observer to allocator was distorted. Second, only the messages concerning a building in the middle of a pattern of three small buildings was distorted. Third, the distortion only influenced information concerning the identification and the sector, thus the type of building remained the same. Consequently, it was no longer possible for the allocator to predict an upcoming fire based on the information received by the messages. The observer, however, could still observe the actual status of the buildings and send the allocator, in time, a message about the expected fire in a large building.

Besides scenarios with distorted messages, "non-routine" scenarios were developed as well (only during the experimental task, not during training), in which the newly learned knowledge about patterns was not applicable. The location of the fire could not be predicted based on the pattern. Nevertheless, the prediction with regard to the type of building (factory or a hospital) remained intact.

2.4 Procedure

Subjects were briefly informed of the general outline of the research (introduced simply as team decision-making research). They were told not to speak to each other about the experiment, and an experimenter was always present in situations where team members were together in the same space (e.g., during breaks). Then subjects were randomly allocated to the role of allocator and observer and they were instructed to read the instruction manual supplied by the experimenter. Further, they trained with the fire-fighting task in two training sessions, consisting of 16 scenarios each.

The instruction manual first explained the fire-fighting task in general, followed by instructions specific for the respective roles. The manual also contained a systematic instruction on how to manipulate the interface accompanied by small tasks that had to be carried out by the subjects. Before pattern recognition was introduced and explained there was a training session of 16 scenarios. After this first training session, subjects were asked to continue to read the pattern recognition instructions. These instructions were followed by another training session of 16 scenarios, which incorporated pattern recognition. The subjects were allowed to ask questions at any point during reading. At the end of the break after the last training session, the subjects were instructed on the experimental condition they were assigned to.

After this instruction the experimental session of 16 scenarios started. The subjects were allowed to use the manual during the experimental session. Each scenario was made up of 12 periods of 15 seconds each. Each team was presented with identical scenarios in a fixed order. The first four scenarios consisted of so called "routine scenarios". In these scenarios, the pre-formatted e-mail messages were not distorted and the knowledge concerning the patterns was always usable. The next four scenarios were also routine, but the pre-formatted e-mail messages were distorted. The following four scenarios contained "non-routine" scenarios with undistorted pre-

formatted e-mail messages. The experimental task ended with four non-routine scenarios with distorted messages. The scenarios were presented in a fixed order, to enable the subjects to apply their newly learned knowledge about patterns. If subjects would have been confronted with non-routine scenarios or scenarios with distorted messages from the beginning, subjects would not have been able to apply this knowledge in later scenarios. In those scenarios, the knowledge about patterns would not have contributed to their performance. Because of this, there would have been a possibility that subjects in a later stage of the experimental task no longer used the knowledge about patterns.

During the training, the two members of the team played the same scenarios at the same time. The subject-allocator played with a computer program that simulated observer behavior (e.g., sending messages etcetera) and the subject-observer played with a computer program that simulated allocator behavior. The programs, or "agents" as they were called, displayed ideal observer and allocator behavior, that is, the agents were always in time with the right information. The subjects were informed of this. Subjects were also informed that in the experimental session they would play with their actual teammate. The choice for this technique was made, to assure an equal level of expertise at the end of the training by controlling the teammate's behavior.

2.5 Design

Within subjects condition

Subjects were presented with eight scenarios without distorted e-mail and eight scenarios with distorted e-mail. Subjects were presented with eight routine and eight non-routine scenarios. In this way four "blocks" of scenarios were formed: two blocks of routine scenarios with and without distorted pre-formatted e-mail messages and two blocks of non routine scenarios with and without distorted e-mail messages.

Between subjects conditions

To examine the impact of verbal communication, four experimental conditions were designed:

1. *Unrestricted communication condition*. In addition to sending and receiving pre-formatted e-mail messages, subjects could communicate unrestrictedly both during and between scenarios. Subjects were placed in the same room and communication was made possible face-to-face.
2. *During scenarios communication condition*. In addition to sending and receiving pre-formatted e-mail messages, subjects could communicate unrestrictedly only from period 2 until period 11 of each scenario. Subjects were placed in separate soundproof rooms and communication was only possible via headsets.
3. *Between scenarios communication condition*. In addition to sending and receiving pre-formatted email messages, subjects could communicate unrestrictedly only between the scenarios and during period 12 from the last scenario and period one of the subsequent scenario during the experimental session. Subjects were placed in separate soundproof rooms and communication was only possible via headsets.
4. *Restricted communication condition*. Subjects could communicate only by sending and receiving pre-formatted e-mail messages through the computer system during the experimental session. Subjects were placed in separate soundproof rooms and verbal communication was not possible, not even between scenarios.

2.6 Dependent variables

Performance measures

A task analysis of the fire-fighting paradigm (see Post et al., 1997) provided information concerning the critical periods of the fire-fighting task. This information was used to define the following performance measures of team performance:

1. *Availability*. A measure of whether sufficient resources were pulled back in period 8. This measure determined for every team and in every scenario, how many fire-fighting units were available in the fire station, ready for allocation to the fire in the target building. There were two possibilities: a team could have sufficient or insufficient units available
2. *Allocation*. A measure of whether sufficient resources were allocated in period 10 to either the hospital or the factory. At the beginning of the fire in the target building, it was determined how many fire-fighting units were assigned. Again, there were two possibilities: a team could have sufficient or insufficient fire-fighting units present at the start of the fire.

Observer rating

Verbal communication was recorded on (video)tape. For every team on every scenario, an observer rated the information exchange using a pre-specified scoring scheme. The following nine categories of information exchange were distinguished:

1. *Data exchange*. Factual information exchange about events in the environment and the status of resources. This includes utterances about the status and location of buildings, the need for fire-fighting units at buildings, and their present allocation. For example, "fire in House A, two units necessary", or "three units present at School B".
2. *Domain knowledge*. Utterances about learned facts of the domain (i.e., from the instruction manual). For example, "a hospital is 1000 casualties", "apartment buildings need two fire-fighting units', or "three complexes in sector one, that means that we are looking for a hospital in sector four".
3. *Meta-communication*. Utterances about the management of the exchange of information. For example, "shall I tell you how many fire fighting units we need on every building that I brief you on"?
4. *Task execution*. This category represents factual information about the work team members are carrying out. That is, the actions they perform on a particular moment in the scenario they are engaged in. Utterances may take the form of logging; that is, explicitly telling the other team member what one is doing at that moment. For example, "I am looking for the hospital right now".
5. *Evaluation of current activities*. Evaluative utterances or judgments concerning activities the subjects are currently engaged in or actions just performed For example, "I don't think this was a good move" or "I think we were too late there".
6. *Planning of current activities*. Utterances about intended activities that do not go beyond the present scenario that the team is engaged in. For example,

"If you move this unit from here to there we might be in time to save the hospital".

7. *General task evaluation.* Evaluative utterances that are based on more than just the current activities or the scenario just played (e.g., "I think we are constantly too slow"). Compared to "evaluation current activities" this type of information exchange is based on the whole *task* and of a higher level of abstraction than "evaluation current activities".
8. *General task strategy.* Utterances that expressed intentions to adjust the way the team should engage in the task in general, deliberations about alternatives, rationalizations of the strategy adopted so far, etcetera. For example, "Okay, from now on we refrain from sending units to houses. We only tackle apartment buildings and schools if they start burning at the beginning of the scenario and we just wait for the big buildings to start burning". Again, this type of information exchange is based on the whole task and is of a higher level of abstraction than "planning current activities".
9. *Remaining information exchange.* Sometimes it was impossible to categorize utterances in one of the categories outlined above. For example, because they were unclear or social in nature.

3 RESULTS

3.1 Performance Measures

On both performance measures ("availability" and "allocation"), teams could score either sufficient or insufficient. Therefore, we performed a non-parametric test. For each comparison, a log-linear model was fitted to the data. The log-linear models we were interested in, contained the two-way interaction of the variables "condition" and "performance" and the three-way interaction of the variables "condition", "performance", and "block".

Period 8: "availability"

The performance variable "availability" assessed the assembling of fire-fighting units in period 8 before the target building started to burn.

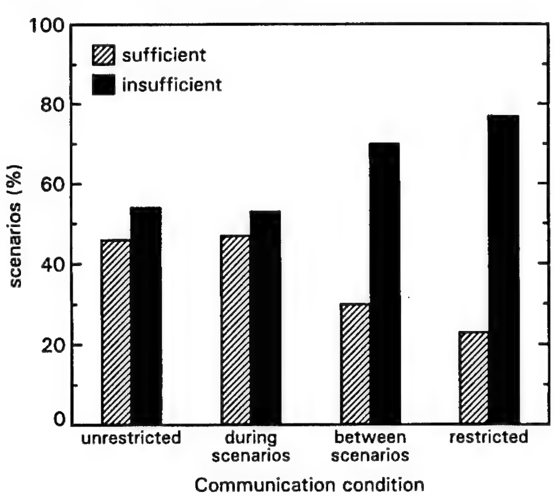


Figure 2 "availability"

Figure 2 shows the percentage of scenarios where a team had sufficient and insufficient resources available during period 8 for each condition. The overall effect for the two-way interaction "condition" and "performance" was significant, $\chi^2(27) = 66.53$, $p < .001$. A post-hoc analysis of paired comparisons of the conditions was performed for "availability". The following five pairwise comparisons yielded significant results:

1. *Unrestricted versus between scenarios communication.* Teams that communicated unrestrictedly assembled sufficient fire-fighting units in more scenarios (46%) than teams that communicated between scenarios (30%), $\chi^2(13) = 36.64$, $p < .001$.
2. *Unrestricted versus restricted communication.* Teams that communicated unrestrictedly assembled sufficient fire-fighting units in more scenarios (46%) than the restricted communication teams (23%), $\chi^2(13) = 36.28$, $p < .001$.
3. *During scenarios versus between scenarios communication.* Teams that communicated during scenarios assembled sufficient fire-fighting units in more scenarios (47%) than teams that communicated between scenarios (30%), $\chi^2(13) = 29.21$, $p = .006$.
4. *During scenarios versus restricted communication.* Teams in the during scenarios communication condition assembled sufficient fire-fighting units in more scenarios (47%) than teams of the restricted communication condition (23%), $\chi^2(13) = 29.15$, $p = .006$.
5. *Between scenarios versus restricted communication.* Teams in the between scenarios communication condition assembled sufficient fire-fighting units in more scenarios (30%) than teams of the restricted communication condition (23%), $\chi^2(13) = 23.24$, $p = .04$.

Figure 3 shows the percentage of scenarios for each "block" where a team had sufficient resources available during period 8 for each condition.

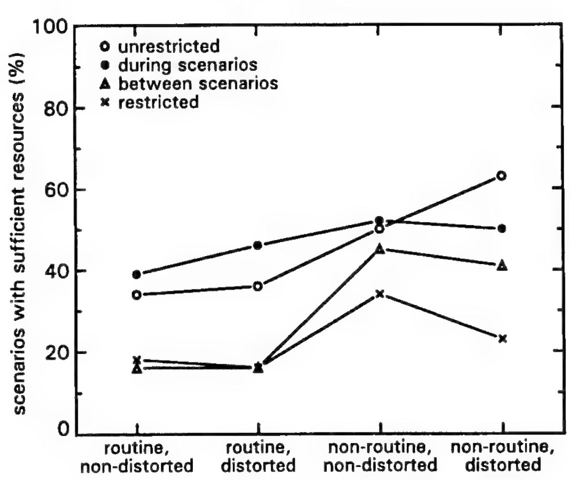


Figure 3 "availability" per block

The overall effect for the three-way interaction "condition", "performance" and "block" was not significant, $\chi^2(24) = 33.09$, $p = .10$. A post-hoc analysis of paired comparisons of the conditions was performed for "availability". The following two pairwise comparisons yielded significant results:

1. *Unrestricted versus between scenarios communication.* Teams that communicated

unrestrictedly were less influenced by the differences in the type of scenario for having sufficient fire units available then the teams that communicated between scenarios, $\chi^2(12) = 26.42$, $p = .009$. Besides that, the teams in the unrestricted communication condition continued to improve their performance, whereas the teams in the between scenarios communication condition were especially influenced by the distorted messages scenarios.

2. *Between scenarios versus restricted communication.* Teams in the restricted communication condition were more influenced by the distorted messages scenarios than the teams in the between scenarios communication condition. $\chi^2(12) = 21.12$, $p = .05$.

Period 10: "allocation"

The variable "allocation" measured whether sufficient or insufficient resources were available at the location of the target building at the start of period 10.

Figure 4 shows the percentage of scenarios where a team had sufficient and insufficient resources allocated during period 10 for each condition.

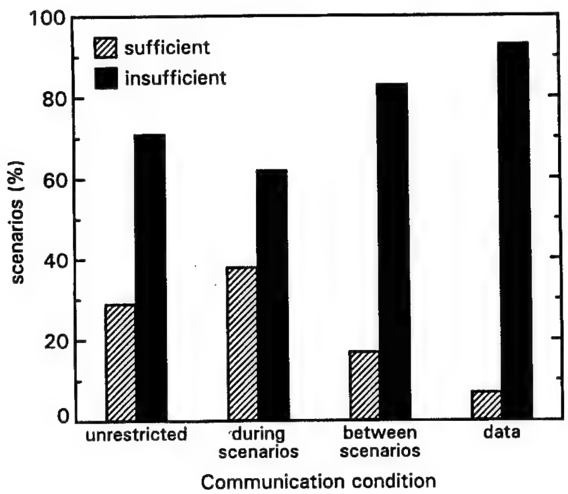


Figure 4 "allocation"

The overall effect of the two way interaction "condition" and "performance" was significant, χ^2 (27) = 87.81, $p < .001$. A post-hoc analysis of paired comparisons of the conditions was performed for "allocation". The following five pairwise comparisons gave significant results:

1. *Unrestricted versus between scenarios communication.* Teams in the unrestricted communication condition allocated sufficient fire-fighting units in more scenarios (29%) than teams in the between scenarios condition (17%), $\chi^2(13) = 35.52$, $p = .002$.
2. *Unrestricted versus restricted communication.* Teams in the unrestricted communication condition allocated sufficient fire-fighting units in more scenarios (29%) than teams in the restricted communication condition (7%), $\chi^2(13) = 38.99$, $p < .001$.
3. *During scenarios versus between scenarios communication.* Teams that communicated during scenarios allocated sufficient fire-fighting units in more scenarios (38%) than teams that communicated between scenarios (7%), $\chi^2(13) = 39.92$, $p < .001$.
4. *During scenarios versus restricted communication.* Teams that communicated during scenarios allocated sufficient fire-fighting units in more scenarios (38%) than teams in the restricted communication condition (7%), $\chi^2(13) = 55.21$, $p < .001$.
5. *Between scenarios versus restricted communication.* Teams that communicated verbally between scenarios allocated sufficient fire-fighting units in more scenarios (17%) than teams in the restricted communication condition (7%), $\chi^2(13) = 28.13$, $p = .009$.

Figure 5 shows the percentage scenarios for each "block" where a team had sufficient resources allocated during period 10 for each condition.

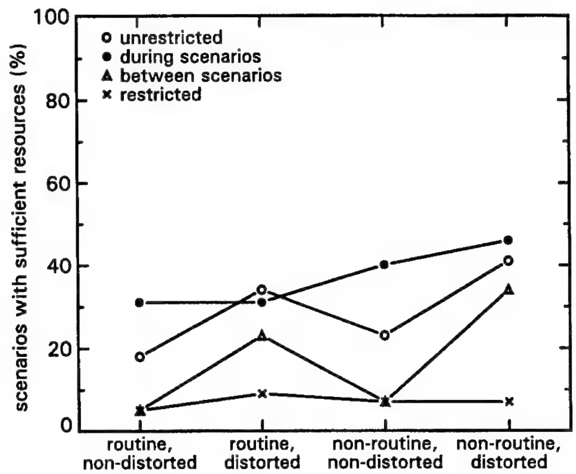


Figure 5 "allocation" per block

The overall effect for the three-way interaction "condition", "performance" and "block" was not significant, χ^2 (24) = 28.73, $p = .23$. A post-hoc analysis of paired comparisons of the conditions was performed for "availability". The following two pairwise comparisons yielded significant results:

1. *During versus between scenarios communication.* Teams that communicated during scenarios were less influenced by the differences in the type of scenario compared to the teams that communicated between scenarios, $\chi^2(12) = 21.00$, $p = .05$. The teams that communicated between scenarios were heavily influenced as a result of the non-routine scenarios, whereas the teams in the during scenarios communication condition were not influenced by the non-routine scenarios.
2. *Between scenarios versus restricted communication.* Teams in the restricted communication condition were more influenced by the distorted message scenarios compared to the teams in the between scenarios communication condition, $\chi^2(12) = 19.12$, p = 09.

3.2 Observer rating

The communication that took place was rated into the different categories described in section 2.5. Table I shows an overview of the total, mean and standard deviation of the utterances for each category aggregated over scenarios.

Table I

	Unrestricted communication			During scenarios communication			Between scenarios communication		
	Total	**Mean**	**SD**	**Total**	**Mean**	**SD**	**Total**	**Mean**	**SD**
Data exchange	2226	139.0	7.0	2176	136.0	4.0	349	21.8	1.6
Task execution information	154	9.6	0.6	371	23.2	1.1	64	4.0	0.4
Evaluation of current activities	337	21.1	1.7	195	12.2	1.3	645	40.3	2.6
Planning of current activities	531	33.2	2.7	299	18.7	2.0	18	1.1	0.3
General task evaluation	219	13.7	1.6	228	14.3	1.8	560	35.0	2.2
General task strategy	206	12.9	1.7	170	10.6	1.8	593	37.1	3.0
Domain knowledge	521	32.6	1.9	458	28.6	1.6	282	17.6	2.3
Meta-communication	26	1.6	0.6	58	3.63	0.8	4	0.3	0.2
Remaining information exchange	296	18.5	1.6	196	12.3	1.2	578	36.1	2.7
Total	4516	282	10.9	4151	259	7.8	3093	193	3.2

Inspection of Table I shows that teams that could exchange information only during scenarios engaged mostly in "activity-based" communication. That is, those teams exchanged information concerning "task execution" and the "planning of current activities" more frequently than teams that could only exchange information between scenarios. The "evaluation of current activities" took place more frequently between scenarios than during scenarios. For the teams that communicated unrestrictedly between the scenarios, utterances that referred directly to the scenario just played were scored under "evaluation of current activities". It is important to note, however, that none of these utterances had the character of abstraction or aggregation of experiences beyond activities in the scenario just played. Finally, teams that could exchange information only between scenarios engaged mostly in "task-related" communication. That is, those teams exchanged information concerning "general task evaluation" and the "general task strategy" more frequently than teams that could only exchange information during scenarios, or teams that could exchange information unrestrictedly all the time.

4 DISCUSSION

Taking the results together, what can we conclude concerning our hypotheses, formulated in the introduction? Our first hypothesis stated that Command & Control teams, that can exchange information unrestrictedly, perform better than teams that can exchange the necessary data only. The results show that when teams have possibilities for exchanging information without restrictions (either between scenarios, during scenarios, or both) their performance increased significantly. Bearing in mind that the teams in each condition could exchange all necessary data in time, we conclude that, for an improved performance, teams need to exchange additional information also.

Our second hypothesis stated that there is a difference in the performance amongst teams, depending on the period during which they can exchange information unrestrictedly. We assumed that teams that could exchange information unrestrictedly only during task execution would engage mostly in "activity-based" communication, whereas teams that could exchange information unrestrictedly between task execution would engage mostly in "task-related" communication. The results show that this is indeed the case. Furthermore, the results show that unrestricted information exchange during task execution contributes significantly more to the team performance when compared to unrestricted information exchange between task execution.

The analysis of the observer ratings of the verbal communication reveal that (besides data exchange), there are several important functions of information exchange that contributed to the team performance. First, exchanging information gives teams extra possibilities for coordination. Teams that have possibilities to exchange information unrestrictedly during the scenarios engaged mainly in "activity-based" information exchange in which team members informed each other what they are doing at that particular moment and what their intentions are to do next. This information allows team members to develop an understanding of each other's tasks and informational needs, which enables team members to provide each other relevant information without explicit request. Second, "activity-based" communication allows team members to detect and correct errors before their consequences become real. Because team members inform each other constantly about what they are doing, fellow team members can respond immediately when things go wrong. Third, exchanging information enables team members to create and maintain a shared understanding of the team, the task, and the situation. A shared understanding of the team gives team members possibilities to clarify each other's roles and capabilities, that allow team members to act upon. In addition, a shared understanding of the situation gives team members the possibility to maintain an accurate understanding of the situation. In a rapidly changing environment, this may enhance team performance, because team members are able to adapt their strategies in time. We expect that "activity-based" as well as "task-related" communication foster the development of a shared understanding. We expect, however, that "activity-based" information exchange fosters the development of a shared understanding of the current activities and an up-to-date view of the ongoing situation, whereas "task-related" communication fosters the development of a shared understanding of the team, the task and the situation in general.

The results support the notion that team self-correction discussions between task execution contribute to the team performance. The analysis of the observer ratings of the verbal communication shows that teams use the time between task execution to engage in evaluating the experience and adjusting their strategies in general for the next time. Information exchange during task execution, however, contributes more to the performance. The analysis of the observer ratings of the verbal communication shows that, during task execution, teams engage mainly in "activity-based" information exchange. When team members do this, they inform each other continuously about what they are currently doing. It is possible that teams have better learning capabilities to improve their performance when team members can provide and receive feedback that is directly linked to a certain action (as it is with "activity-based" information exchange). This result is supported by the notion of the ACT-R theory that feedback must be linked to a specific situation as much as possible.

It is hypothesized that it is difficult for teams after task execution to establish and remember where it went wrong. Especially when teams have to deal with complex circumstances, it can be questioned whether teams are able to comprehend where it went wrong and, know what to do about it.

Finally, it is possible that "task-related" information exchange has limited value when teams encounter novel situations. After all, when situations change unexpectedly, previously learned strategies cannot be used any more. When teams have to deal with novel situations, it is possible that teams have to rely on "activity-based" information exchange, because this allows teams to adjust their strategies in time. The results give some support to this view. That is, teams that engage in "task-related" communication after task performance show an improvement in their performance as they executed several scenarios of the same type. When they have to deal with a new situation, however, their performance declined. After executing several scenarios that contain the "new" situation, they recovered and improved their performance again. The teams that engaged in "activity-based" communication during scenarios, on the other hand, continue to improve their performance after several scenarios, regardless of the unexpected changes. Therefore, we suggest that "activity-based" communication may play an important role when teams have to handle new situations.

4.1 Conclusions

Four conclusions can be drawn from the present results. First, in order to improve the performance in Command & Control teams, members need to exchange information, in addition to the necessary data. Second, these results support the notion that team self-correction contributes to the team performance in so far as teams that engage in "task-related" communication after task execution perform better than teams that exchange the necessary data only. Third, "activity-based" communication during task execution contributes more to the team performance than "task-related" communication after task execution. This can be explained by assuming that this type of communication leads to better coordination, cross check and correcting of errors, development and maintenance of a shared situation understanding, and more possibilities to learn. Fourth, "activity-based" based information exchange is required to deal successfully with novel situations.

5 REFERENCES

Adams, M.B. (1995). Functional Analysis/ Decomposition of Closed-loop, Real-time Work Processes. *Proceedings of the AGARD Mission System Panel*, 1-10.

Anderson, J.R. (1993). *Rules of the Mind*. Hillsdale, New Jersey, Lawrence Erlbaum Associates.

Blickensderfer, E.L., Cannon-Bowers, J.A., & Salas, E. (1994). Fostering Shared Mental Models through Team Self-correction: Theoretical Bases and Propositions. In D. Johnson & S. Beyerlein (Eds.). *Advances in interdisciplinary studies in team work series*, vol. 4, Greenwich, CT: JAI Press Inc.

Cannon-Bowers, J.A., Tannenbaum, E.S., Salas, E., & Volpe C.E. (1995). Defining Competencies and Establishing Team Training Requirements. In R.A. Guzzo & E. Salas (Eds.). *Team Effectiveness and Decision-making in Organizations* (pp. 333-379). San Fransisco, Jossey-Bass Inc.

Dyer, J.L. (1984). Team Research and Team Training: a state of the art review. In F.A. Muckler (Ed.), *Human Factors Review* (pp. 285-323). Santa Monica: The Human Factors Society.

Hutchins, E. (1992). The technology of team navigation. In J. Galegher, R. E. Kraut, & C. Egido (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work* (pp. 191-220). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

McIntyre, R. M., & Salas, E. (1995). Measuring and Managing for Team Performance: Emerging Principles from Complex Environments. In R.A. Guzzo & E. Salas (Eds.), Team Effectiveness and Team Decision-making in Organizations (pp. 9-45). San Francisco, Jossey-Bass Inc.

Orasanu, J., Salas, E. (1993). Team decision-making in complex environments. In G.A. Klein, J. Orasanu, R. Calderwood & C. E. Zsambok (Eds.), *Decision-making in action: models and methods* (pp. 158-171). Norwood, NY: Ablex Publishing Corporation.

Post, W.M., Rasker, P.C., & Schraagen, J.M.C. (1997). *The role of communication and coordination in team decision-making in a command & control task.* (Report TNO-TM 97-B004). Soesterberg NL: TNO Human Factors Research Institute.

Passenier P.O., & van Delft, J.H. (1995) De mens-machine interface [The Human-Machine Interface (in Dutch)]. In van Delft, J.H., & Schuffel, H. (Eds.) (1995). *Human Factors onderzoek voor toekomstige Commando Centrale KM* (pp. 51-58) [Human Factors Research for future operations rooms Royal Dutch Navy (in Dutch)] (Report TNO-TM 1995 A-19), Soesterberg NL: TNO Human Factors Research Institute.

Rasmussen, J. (1993). Deciding an Doing: Decision-making in Natural Contexts. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.) *Decision-making in Action: models and methods* (pp. 158-171). Norwood, NJ: Ablex Publishing Corporation.

Salas, E., Dickenson, T.L., Converse, S.A., & Tannenbaum, S.I. (1992). Toward a Understanding of Team Performance and Training. In W. Swezey, & E. Salas (Eds.). *Teams: their Training and Performance* (pp. 3-29). Norwood, NJ: Ablex Publishing Corporation.

Schraagen, J.M.C. (1995). *Besluitvorming in teams: een onderzoeksparadigma* [Team decision-making: a research paradigm (in Dutch)]. (Report TNO-TM 95 B-1). Soesterberg NL: TNO Human Factors Research Institute.

Seifert, C.M., & Hutchins, E.L., (1992). Error as opportunity: Learning in a cooperative task. *Human Computer Interaction*, 7, 409-435.

Serfaty, D., Entin, E.E, & Volpe, C. (1993). Adaptation to Stress in Team Decision-making and Coordination. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, 1228-1232.

Stout, R., & Salas, E. (1993). The Role of Planning in Coordinated Team Decision-making: Implications for Training. *Proceedings of the Human Factors Society 37^{th} annual meeting*, 1238-1242.

Van Delft, J.H., & Schuffel, H. (Eds.) (1995). *Human Factors onderzoek voor toekomstige Commando Centrale KM* [Human Factors Research for future operations rooms Royal Dutch Navy (in Dutch)] (Report TNO-TM 1995 A-19), Soesterberg NL: TNO Human Factors Research Institute.

The Impact of Environment Complexity and Ambiguity on AWACS Crew Performance in Simulated Combat Scenarios

M. Glenn Cobb, Lt. Col., USAF
Department of Psychology
The Pennsylvania State University
417 Bruce V. Moore Bldg.
University Park, PA, USA 16802-3104

John E. Mathieu, Ph.D.
Linda R. Elliott, Ph.D.
Mathieu A. Dalrymple

SUMMARY

Using data gathered in highly demanding combat simulations by the Aircrew Evaluation Sustained Operations Performance (AESOP) facility, Brooks AFB, TX, we examined how the interaction of environment complexity and cue ambiguity affected team performance by impacting intervening variables such as team communication, coordination, and situational awareness. Five operational Airborne Warning and Control System (AWACS) teams each completed five scenario-based mission simulations. Each scenario represented a dynamic continual flow of activity, but was easily partitioned into four essentially equivalent waves of action. Within each wave were embedded six specific decision events that were designed to elicit team interaction and decision making. Evaluations of each team's decisions were made by operational subject-matter experts (SMEs) based on an a priori listing of appropriate team responses. We recognized that such comparisons may provide an incomplete picture of overall team effectiveness due to the impact of earlier decisions on the resources available for later events. To better capture the potential impact of this effect, SMEs' ratings also took into account the situation that the crews actually faced, regardless if it resulted from scenario design or from earlier crew actions and decisions. Team process was expected to play a key mediational role. Specifically, environmental complexity and cue ambiguity were expected to impact overall team performance outcomes by influencing the effectiveness of team decision making process strategies and behaviors. Findings are discussed in terms of their implications for team training in dynamic environments. The original data collection was conducted in a research project sponsored by the Air Force Office of Scientific Research.

1. INTRODUCTION

Critical to achieving information and battlespace dominance within any area of military operation is the ability to maintain effective command and control of all combat resources (Ref 1; Ref 2; Ref 3). Worldwide reductions in military forces and operational military bases have greatly increased our dependence on highly mobile, airborne platforms to effectively orchestrate allied resources in the combat arena. Enhancing our understanding of the influences affecting the effectiveness of deployed command, control, and communications (C^3) teams, especially US Air Force (USAF) AWACS crews, has become

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

an increasingly important concern for military commanders and mission planners. In the USAF, the E-3 AWACS provides all-altitude surveillance of airborne assets and hostile targets over land and sea. AWACS missions involve detecting enemy aircraft, controlling defensive friendly fighters, controlling ground strike aircraft, and providing a long-range picture of airborne activity in a region to theater commanders and other command staffs. AWACS aircraft carry no armament and a flight crew of four with 19 to 29 mission specialists (Ref 1).

Each of the US military services have constructed specialized laboratories dedicated to improving our operational warfighting and force management capabilities by extensively examining the interrelationships between a dynamic warfighting environment and those personnel required to orchestrate the implements of tactical warfare. The intellectual and analytical resources of civilian universities in the United States are also being called upon to assist in this research effort. This study marks one such collaborative effort between researchers and technicians at the Air Force Research Laboratory, at Brooks AFB, TX, and researchers at the Pennsylvania State University specializing in team research and analysis.

Using archival data collected by USAF researchers, this research examined how the interaction of environment complexity and cue ambiguity affected team performance by impacting intervening variables such as team communication, coordination, and situational awareness. This is an especially important consideration in military command and control situations, since the consequences of decision errors can be devastating in terms of battle outcomes and loss of lives. This effort continued the research focus suggested by Weaver, Bowers, Salas, and Cannon-Bowers (Ref 4), who asserted that it is crucial to examine team performance under multi-level, multi-task conditions. This research contributes to the existing team performance literature by analyzing the factors and processes influencing experienced crew performance under highly realistic simulation conditions, task demands, and operational performance standards.

Given that numerous definitions of "teams" exist in the literature (Ref 5; Ref 6), we believe Salas, Dickinson, Converse, and Tannenbaum's (Ref 7) definition best describes the general type of operational teams found in military organizations. They defined a team as "...a distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal/objective/mission, who have each been assigned specific roles or functions to perform, and who have a limited life-span of membership" (p.4). Central to this definition is that effective performance requires a dynamic exchange of information and resources among the team members, coordination of task activities, constant adjustments to task demands, and organizational structuring of team members (Ref 7). Since much of the research referenced within this paper has treated "groups," especially work groups, and "teams" as interchangeable terms, we will simplify matters by using the term "team" for all cases where Salas et al.'s (Ref 7) definition applies.

2. MODEL

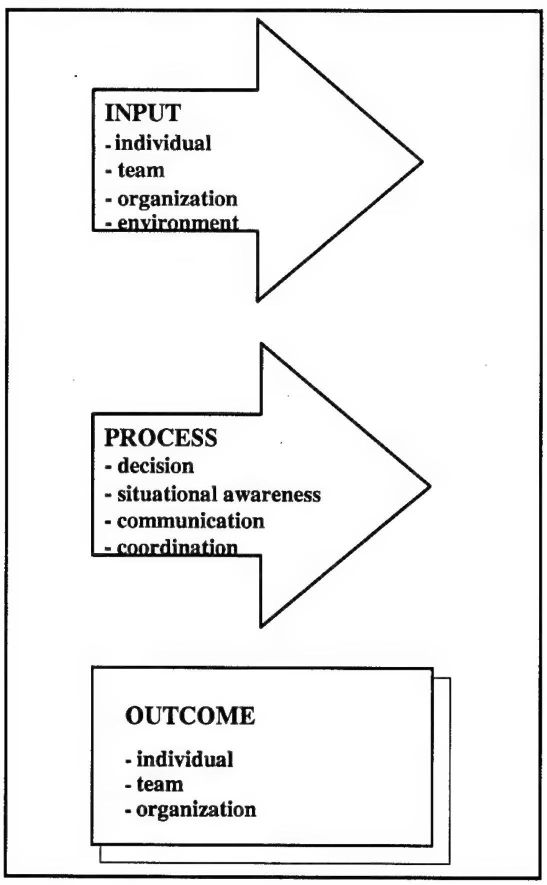
Whether the actors involved are individuals, teams, or organizations, understanding how various stimuli, information, or other inputs affect decision making processes is central to understanding how these factors influence performance outcomes. To make sound decisions that result in beneficial outcomes, actors rely on their understanding and perception of the situation around them. When the situation provides inadequate or conflicting information, actors generally attempt to gain additional information or insight by communicating with others and by actively exploring the situation in greater depth.

Additionally, the environment can impact this process by providing multiple and sometimes conflicting sources of information, and by placing multiple task demands upon an actor. The more things an actor has to deal with or act on, the more complex the task of making an accurate and effective decision. In this research, we used a general model to illustrate and examine these relationships within an AWACS crew reacting to the demands of a regional conflict simulation.

2.1. *The I-P-O Model.*

In general, most current research examining team performance and effectiveness has followed Hackman and his colleagues' inputs-process-outcomes (I-P-O) framework (Ref 8; Ref 9; Ref 10; Ref 11) as illustrated in Figure 1.

Figure 1: Team Process Model



Outcomes are the products of team activity that are valued by either the organization or the team members. These outcomes characterize the effectiveness of the team and generally focus on quality, quantity, and timeliness of the products the team produces (Ref 11). With the exception of the initial formation of the team, the outcomes from one performance episode will influence the inputs of the next performance episode. In a dynamic environment, such as air combat operations, teams are usually required to produce multiple simultaneous outcomes. Team members are also often required to simultaneously perform individual tasks while fulfilling team responsibilities. For example, AWACS crew members are directly responsible for simultaneously tracking and assisting many different types of aircraft, while ensuring the team is fulfilling critical objectives established by senior commanders. The crew and each of its members need to remain very conscious of the exact outcome expectations of the various constituencies they are supporting. In combat environments, most decision outcomes must be high in quality to prevent or minimize losses among friendly forces while meeting mission objectives.

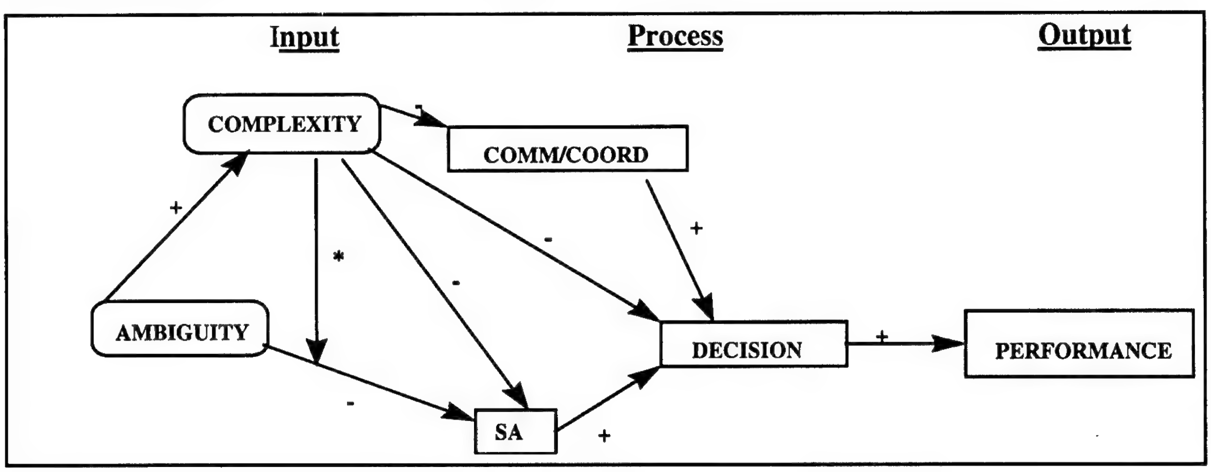
Process factors describe how team members interact and collectively deal with their environment. They include variables such as communication, cooperation, coordination, cohesion, leadership, situational awareness, and decision making. Process mediates a team's effectiveness in creating desired outcomes from various inputs. Even under optimal environmental conditions, poor team process can lead to process loss (Ref 12) which results in less than optimal performance. Conversely, in suboptimal operating conditions, good team processes may enable a higher team performance outcome than might otherwise be expected considering the situation.

Inputs include conditions that exist prior to a performance episode, such as team composition, task demands, member characteristics, and

organizational and contextual influences. Performance episodes are "distinguishable periods of time over which performance accrues and is reviewed" (Ref 13, p. 1759). An inherent assumption in the I-P-O model is that there are separate inputs, processes, and outcomes for each performance episode. As Hackman (Ref 9) pointed out, individual members of a team provide an additional, and important, contextual input for a single member's behavior and reactions within the team. In their review of networked computer simulations designed to explore team performance, Weaver et al. (Ref 4) noted that teams are frequently hierarchical and have varying degrees of responsibilities and expertise among their members. This is especially true in military organizations where the rank and experience level of individual crew members may be very different within the same team. A second lieutenant, fresh out of college and technical training, could suddenly be placed in a responsible position, directing a crew of master sergeants with nearly 20 years of task experience. This type of composition requires that a team establish a balance between official positions of authority and sources of task expertise critical to achieving mission requirements and desired performance outcomes. Balancing task requirements with the knowledge, skills, and attitudes (KSAs) present within the team to produce specific outcomes is an important aspect of team process.

In the current effort, we anticipated that performance outcomes would be directly influenced by the effectiveness of team decision making. In turn, team decisions would be impacted by how well teams recognized and interpreted the situational demands surrounding them, and how well they coordinated information and task responsibilities among individual team members. We expected the complexity of the task environment, especially as indicated by team activity levels, and the ambiguity of critical decision cues within the environment to affect the effectiveness of the teams' communication, coordination, and situational awareness (see Figure 2).

Figure 2: Hypothesized Team Performance Model



2.2. *Team Performance.*

The effectiveness of a team's performance reflects how well a team accomplishes its purpose or mission (Ref 14). Although, as illustrated by the many variations present in the literature, effectiveness can be quantified differently for almost every type of team or context, almost all of the variations are concerned with evaluations of a team's performance. Research and organizational assessments have routinely focused on the quantity and quality of the services and products provided by the team. Interestingly, Tannenbaum, et al. (Ref 14) also included a team's ability to overcome uncertainty, "to remain vital and alive and to grow and regenerate itself," in their view of team effectiveness (p.505). They emphasized that this critical characteristic enables a team to sustain its performance and fulfill its purpose over a period of time despite dynamic and changing task environments.

Kozlowski and Gully (Ref 15) asserted that most training programs are based on making the right decision to achieve some specified outcome. However, they challenged this notion by emphasizing that rational analysis does not always '... yield optimal solutions when the problem is ill-defined, information is ambiguous, and the situation is dynamic" (p.4). Their focus was directed to assessing the appropriateness of the team's overall decision process, i.e., making the decision in the right way. Although they concentrated on the adaptive expertise of teams from a learning perspective, Kozlowski and Gully made a very valid point. Real-world situations and problems rarely fit rational models of outcome-based decision models. While a strength of their research was in linking active learning processes with adaptive expertise, Kozlowski and Gully, as well as other researchers, still seemed to ignore how team decisions can radically change their environmental context. Put another way, environments are not only dynamic and fluid, they are continually evolving as well. In the current effort, we anticipate that the effectiveness of teams' decision making processes will directly affect the teams' performance outcomes.

2.3. *Team Process.*

All teams must effectively and collectively process the various inputs they receive and deal with interpersonal behaviors and conflicts to efficiently produce desired or directed outcomes (Ref 9). In their review of research on team effectiveness, Guzzo and Dickson (Ref 8) contended that military flight crews are very different from other aviation crews. From a military perspective, C^3 teams, like AWACS crews, have the same general attributes as flight crews. Not only do they encounter equally complex and demanding task environments, but their training is designed along the same general lines as most flight crews. Military flight crews are likely to remain together for longer periods of time, have a formal rank structure, and fly a greater number of training flights than their civilian counterparts (Ref 8). Yet, even with these advantages over civilian crews, research shows that military crews also experience similar problems with information exchange, crew relationships, and task prioritization (Ref 8; Ref 15). Military organizations have spent a great amount of resources and time to developing training programs and standard operating procedures to improve team performance by enhancing effective team process.

In summarizing the existing research on team training, Salas, et al. (Ref 7) pointed out that informing team members of the nature and requirements of other team members' subtasks is an effective way to emphasize the need for communication and coordination. This type of training and expertise encourages team members to identify their interdependencies, recognize when and to where information must be transferred, and understand the consequences

of failing to effectively coordinate their respective task responsibilities and decisions. They also asserted that existing research suggested that effective teams under high stress conditions employ implicit coordinating mechanisms. These mechanisms allow team members to anticipate others' needs while reducing the need for overt communication. When the unexpected occurs, or there are no standard procedures to follow, or a team member does not have the expertise needed for the task, interaction with other team members becomes essential for making effective decisions.

The ability to identify and remain cognizant of changes in the environment and relevant task demands is a critical performance component of highly dynamic and complex tasks (Ref 17). Situational awareness includes perceiving current environmental elements in both time and space, comprehending the meaning and significance of these elements, and predicting future actions of these elements in the immediate future (Ref 18). Research and experience have repeatedly shown that situational awareness is fundamental to successful team performance in military, organizational, and industrial settings. Kozlowski and Gully (Ref 15) estimated that 80% of aircraft accidents, as well as tragic events like the destruction of a civilian airliner by the USS Vincennes in the Persian Gulf, are attributable to errors made by highly trained, experienced specialists operating in a team context. Official investigators and academic researchers have tracked many of these errors to a breakdown in a team's ability to recognize and adapt to unexpected changes in the environment. Although it is frequently studied by military researchers in terms of aircraft cockpit designs and head-up displays (Ref 19), situational awareness in team process has gained in importance. For example, enhancing individual and collective situational awareness has become increasingly significant in pilot and aircrew training programs (Ref 20). Yet, as Mathieu, Goodwin, Heffner, and Born (Ref 21) pointed out, there has been relatively limited research focusing on the role of contextual influences on team processes. While some researchers have emphasized the importance of accounting for the impact of organizational context or environment on team effectiveness (Ref 22; Ref 7), Mathieu et al. noted that even among the few studies empirically examining this issue, teams are seen as largely passive respondents to situational demands.

Following Ancona and Caldwell's research indicating that teams are affected by and, in turn, affect their environment (Ref 23; Ref 24), Mathieu et al. (Ref 21) examined the hypothesis that team processes and environmental demands co-evolve over time. They determined that as teams choose task strategies and execute decisions in response to diverse situational demands, their actions alter the resources available for future actions and, to some degree, the environment surrounding them. For example, an AWACS crew vectoring friendly resources to respond to an enemy threat not only changes the friendly resources available for future responses, but the success or failure of their decision also impacts possible reactions by the adversary. Thus, an action-reaction cycle begins that directly impacts the potential complexity of future situations and the very nature of any further threat options.

By consistently relying only on comparisons with predetermined "true scores," researchers may be overlooking the impact team decisions have on their planned course of events. We contend that it is necessary to also evaluate team performance relative to the actual context confronting the team when the decision is made. A teams' initial decisions can dramatically change the context of subsequent decisions and behaviors by altering the resources available to support later decisions and strategy options. The situation becomes even more complex, when one considers the fact that teams rarely act in total isolation in "real-world" environments.

One team's outcomes and decisions become the stimuli or inputs for other teams and individuals working in related or competing areas of responsibility. For example, an AWACS crew's decisions directly affects the actions, decisions, and risks faced by the pilots to whom they are providing information and threat assessments. Their team decisions can also alter the anticipated reactions of a hostile adversary, thus fundamentally changing the threats encountered by all the combat teams, including themselves. Thus, by recognizing, understanding, and adjusting to the impact of actual C^3 decisions by AWACS and other teams, commanders can adjust their strategies to reduce uncertainty and maintain order in the face of emerging chaos. Based on this research, we hypothesized that the quality of a team's communication, coordination, and situational awareness would collectively impact the effectiveness of the team's decision making processes, as indexed by SME ratings.

2.4. *Environment Complexity.*

To varying degrees, most models of team effectiveness (Ref 22; Ref 25; Ref 14) address the impact of organizational and situational factors, task demands, team member attributes, and team composition on performance outcomes. The greater the need for team members to depend on and directly support one another to accomplish tasks, the more complex their interdependence and environment will be (Ref 26). If a specific team member has all the resources, information, authority, and means needed to make and carry out a decision, there is little reason to interact with other team members. This person's task is greatly simplified, since time and cognitive resources do not have to be devoted to ensuring a teammate understands the situation, agrees to the intended response, and provides the resources or assistance needed to fulfill the task.

Salas et al. (Ref 7) emphasized that complexity can be measured by many different metrics, including objective and subjective measures of information processing demands, time constraints, and workload. They asserted that no single metric applies to all types of teams and that multiple metrics should be used to analyze team tasks. Salas et al. pointed out that task-oriented team performance models emphasize that teams require very little communication to be successful when tasks have low interdependency, low complexity, and a hierarchical work structure. Yet when the opposite is true, i.e., high interdependencies, significant complexity, and a decentralized work structure, team members must communicate frequently and effectively to make effective decisions. For AWACS crews, as for many other military and civilian teams, the timely and accurate transfer of information is absolutely critical to effective performance. Therefore, we hypothesized that increasing levels of environmental complexity would negatively impact the quality of a team's communication and coordination process by limiting the amount of time and resources available to confirm information transfers, clarify task responsibilities, and anticipate environmental changes.

2.5. *Cue Ambiguity.*

A task strategy's effectiveness depends on the specific requirements of a task that can be identified by team members based on contextual cues present in the environment and within the team's actions (Ref 9; Ref 10). Hackman and Oldham (Ref 10) insisted that team members must first assess task requirements and constraints, then develop strategies based on these perceptions. These strategies become associated with certain environmental attributes and are initiated when specific decision cues are detected in the environment by team members. Hackman (Ref 9) also emphasized the important role played by interpersonal cues within the team that speed up collective decision making and information processing. Team members develop behavioral cues that accelerate problem

solving and verifying decision consensus. Recognizing cueing events or stimuli, in the environment and within the team, is an important part of military training in field exercises and operational simulations.

Over the course of time and with repeated practice, specific status indications or events in the weapon system (such as an emergency audio signal) or the task environment (such as an enemy's movement of forces) become associated with specific crew actions and environmental indications. To enhance a common understanding of the situation, military personnel use specific audio cues to act as a verbal shorthand for relaying important evaluative judgments of the quality of the situational information they are reporting. For example, a fighter aircraft flies over a previously unknown group of enemy forces. If the pilot is uncertain of the nature of the forces, he will caveat his report to the AWACS crew as either an estimate or a probable sighting. This additional information helps the AWACS crew assess the nature of the threat and determine the most appropriate course of action. For example, if the original pilot was uncertain about the sighting and the force potentially represented an important threat, the AWACS crew could send another flight over the site to check things out more thoroughly rather than immediately committing resources to destroy the new target. The goal in this type of training is to ensure crews can react effectively and efficiently, in terms of time and resource allocation, to changes in their task environment. In operational training and evaluation simulations and procedures, decision cues are specifically designed and selected based on their capacity to draw attention to themselves. A major problem in complex environments is that there are many decision stimuli vying for a crew member's and the team's attention. Information processing theories indicate that people simply cannot actively attend to all simultaneously occurring stimuli equally (Ref 27). The perceptual analysis of unattended information is attenuated or reduced. A personally relevant or significant stimulus can allow weakly attended stimuli to cause a shift in attention. An example of this is the well-known cocktail party effect in which an individual can sometimes hear their name in a crowded room of party-goers (Ref 27). However, there is a limit to such effects as the environment can become so noisy, or complex, that it simply is no longer possible to detect or isolate a specific, weakly sent cue amid all the other stimuli. To enhance a crew's skills at recognizing cues and matching them to required decisions, it is standard practice to manipulate the clarity of visual and auditory cues during military training simulations. Various techniques, such as adjusting the quality of an audio message or embedding the cue within a multiple task sequence, are also used to increase and decrease a cue's ambiguity. Therefore, we hypothesized that cue ambiguity impacts teams' decision making process by affecting their situational awareness.

3. METHOD

3.1. *Facility: Air Force Research Laboratory.*

The data supporting this research was collected in the AESOP facility at the Air Force Research Laboratory-Armstrong Research Site (AFRL-ARS) at Brooks Air Force Base, TX, from October 1996 through June 1997. The AESOP facility offers researchers a unique opportunity to examine the complex decision making behaviors occurring within highly trained, experienced, AWACS teams operating in a realistic simulation. Personnel from the AFRL-Wright Research Site (AFRL-WRS) at Wright-Patterson AFB, OH, designed and constructed four generic workstations, called the Command, Control, and Communications Simulation Training and Research Systems (C^3STARS), at AFRL-ARS to replicate the weapons director (WD) crew stations of an AWACS platform (Ref 28). These four stations employ high-resolution graphic displays, modular switch panels with programmable switch functions, communication panels, keyboards, and

trackballs. The design of the system not only allows researchers to make audio and video records of team performance, but it has the capability to capture the frequency of each crew member manipulation of the equipment (e.g. switch actions, foot presses, and screen enlargement) (Ref 28). The system has also been expanded by connecting the generic C^3 consoles to the Defense Simulation Internet (DSI), thus allowing integration with assets at other Department of Defense (DOD) facilities into a multi-force, joint service simulation exercise employing multiple teams operating at separate locations (Ref 28).

3.2. *AWACS WD Simulator.*

While there are some relatively minor differences between the laboratory workstations and those aboard AWACS aircraft, they do not present a significant obstacle to the experienced crews being examined (Ref 29). Due to the aircraft's greater restrictions in workspace, the crew positions are slightly farther apart in the AESOP facility's simulator. The WD positions are set side-by-side, as on the actual aircraft, to enhance communication and the exchange of information between the WDs. As on the aircraft, the positions are still close enough together in the AESOP simulator to allow one WD to peer at the display screen of the WD seated next to him/her. There are other relatively minor differences in the console layout, but, according to SMEs and participant feedback, these differences are inconsequential and have no impact on the team members' ability to communicate and perform in the same manner as they do aboard the aircraft. Overall, the crew positions closely replicate the set-up found in the aircraft and in operational training simulators. To ensure that crews have time to adjust to the differences between AWACS and AESOP consoles, experience and participant feedback have indicated that a 30- to 40-minute familiarization training session is required. Previous research participants have consistently reported little problem in adjusting to the facility.

3.3. *Scenario Development.*

Scenarios concentrated on simulating the demands of a localized regional conflict where the team was the only C^3 resource in the area. Based on the recommendations of operational training instructors, the new scenarios reduced the average number of total tracks in any given wave from 120 to 60. In this context, this total refers to the sum of all friendly and hostile aircraft being monitored by the AWACS crew. The term track is used in reference to the flight path of the aircraft as indicated by radar returns. A WD's primary function is to track all aircraft operating in the area of responsibility. The research designers had also previously noted an unacceptable response delay in the simulator's data processing. Due to the nature of this research, it was absolutely vital that the simulator replicate real-time responses as accurately as possible. Reducing the number of tracks presented during each wave was also deemed necessary to avoid immediately overwhelming the experience levels of the WDs selected for the study. Finally, SMEs tested the scenarios and their presentation to ensure they followed current AWACS doctrine and recent field experiences.

The resulting data collection continued ongoing research being conducted by USAF investigators exploring AWACS teams' effectiveness under conditions of variable information availability and quality embedded in high-stress, dynamic combat simulations (Ref 28). Based on Fowlkes, Lane, Salas, Franz, and Oser's (Ref 30) research techniques, the personnel at AFRL-ARS developed embedded decision events to focus on specific team behaviors and activities. In the original research design, each of six independent scenarios contained four waves of hostile aircraft activity lasting 30 minutes each, with an approximately 10-minute separation between attacking waves.

Embedded within each wave were six decision events that provided the focal points for data collection and performance comparisons. When analyzing the data, this design structure provides 144 decision events (6 scenarios x 4 waves x 6 decision events) across an N of 8 teams (total of 1152 decision/performance episodes). Since the decision events overlapped with one another and could be delayed by other crew priorities, it was decided that the wave level of analysis provided the most firm basis for examining team process and performance (N = 192 (8 x 6 x 4)). Due to the time required for encoding and analysis, the impact of other research projects, and limited archival performance scores, this paper is based on the analysis of 5 teams across 5 scenarios. This creates N = 100 observations (5 x 5 x 4), when analyzed at the wave level.

To facilitate position interdependencies, research SMEs parceled mission C^3 responsibilities into three functional WD roles: high value asset, combat air patrol, and strike. Elliott et al. (Ref 28) pointed out that this procedure replicates the clarification and distribution of position responsibilities carried out by AWACS crews during mission planning. Crew members were also rotated through the three positions between each scenario to simulate common and accepted mission operations. Since each scenario was designed to be distinct from previous scenarios, the researchers also hoped this positional rotation would enhance the teams' ability to make a clean mental break from one scenario to the next. In this way, task carry-over from one scenario to another would be minimized.

According to the research designers at the AFRL-ARS, the division of labor between these three WD positions duplicated normal mission responsibilities and resources. The high value asset (HVA) WD controls all C^3 aircraft, air refueling operations, electronic warfare (EW) operations, and reconnaissance platforms. These assets are considered to be "high value" because they provide critical support to fighting units. Without effective C^3, it is impossible to orchestrate all the various types of aircraft and combat units within a dynamic warfighting environment. Air refueling is critical to keeping aircraft in the fight and extending their combat fighting range deep into enemy territory. It is simple--without fuel, you do not fly and fight. EW platforms are dedicated to detecting and jamming enemy weapon systems and their attempts to prevent our effective use of the electromagnetic spectrum to detect enemy movements through radar. These actions are vital to maintain our C^3 capabilities, while denying an adversary the same freedom of action. Warfare in the electronic information arena is at the heart of Admiral Owens' (Ref 2) battlespace dominance concept and our information warfare strategies. Success in EW provides our fighting forces with a decided advantage in being able to carry out integrated attacks against confused or isolated enemy forces. Reconnaissance provides vital intelligence and battle damage verification (feedback) to assist commanders in their combat decisions.

The combat air patrol (CAP) WD controls defensive counterair (DCA) aircraft, fire coordination of friendly surface-to-air missile (SAM) assets, and overall team leadership. DCA operations are dedicated to protecting our friendly forces from enemy attack. The AWACS crew detects attacking enemy aircraft and directs defenders to the best positions to intercept and destroy the threat. Due to the high interdependency of their position with the other two WDs, the importance of their responsibilities to overall team success and their central console position, the CAP WD normally provides leadership for the team and helps resolve resource conflicts in the simulation.

The Strike WD controls planned bombing, unplanned SEAD, and unplanned theater missile defense (TMD) bombing missions (Ref 28). The strike position focuses on air-to-ground

attack operations. The aircraft controlled by this position are dedicated to destroying key ground targets in support of campaign objectives. These missions can be planned, based on prebattle analysis and planning, or unplanned, e.g., a previously unknown ground target is discovered during the course of the warfighting. This position is also responsible for coordinating preplanned attacks against enemy ground forces, known as suppression of enemy air defenses (SEAD), designed to destroy incoming aircraft. The most important of these potential targets is the enemy's SAM sites, since they are a threat to any aircraft flying in the area. Destroying or degrading the capabilities of these sites is essential to maintaining our freedom of operations and reducing our aircraft losses to a minimum. This position provides important flexibility to the team due to its focus on reacting to developments in the fighting as they occur. This position also controls aircraft that are capable of conducting both air-to-ground and air-to-air missions, depending on their weapons load. This resource flexibility is very important in a team's ability to react to changes in the environment and unexpected enemy reactions.

During the course of succeeding waves, team members normally discover that they need to hand off aircraft to each other and to keep each other appraised of the evolving situation and positioning of air assets. For example, one of CAP's fighter aircraft is running low on fuel, so he hands the fighter off to HVA for air refueling. While HVA is handling the refueling, CAP requests one of Strike's swing-role, fighter/bomber aircraft to cover the vacated defensive patrol. Strike hands the aircraft off to CAP and concentrates on taking out an enemy SAM site. Destroying the SAM site will allow HVA to reposition his tankers closer to the defensive perimeter, so CAP can reduce the incoming air threat and begin to take the fight to the enemy. As you can see from this brief example, the three WDs are very dependent on each other for coordinating resources and information as they try to handle multiple tasks simultaneously. It also illustrates how teams can use their resources to alter the nature of their environment to their advantage.

3.4. *Participants.*

Eight 3-person teams were comprised of operational WDs from Tinker AFB, OK. These personnel were primarily selected by their commanders based on availability, scheduling, and suitability for the specific research project. According to AESOP personnel, availability seemed to be the primary factor determining a subject's selection. Unfortunately, no female volunteers were available to participate in the research effort. Participants were all experienced, fully qualified WDs. All the participants were USAF personnel with one exception--one Canadian officer was selected to participate in the project. While his training was very similar to his US counterparts, he had to rely on his experience to adjust to working with US personnel procedures. He was also one of the more experienced participants and had flown previous missions with US resources. Biographical data were collected on each subject to determine the nature and level of their weapon system experience. The participants were of various ranks ranging from O-1 to O-4 and E-3 to E-5 (Major: 1, Captains: 6, 1Lts: 2, 2Lt: 1, TSgt: 1, SSgts: 7, SrA: 6). They ranged in age from 23 to 43 years old. Teams were constructed by operations personnel at Tinker AFB and arrived at the same time at AFRL-ARS for the project. Team selections followed standard operating procedures for AWACS crew deployment scheduling.

3.5. *Procedure.*

Replicating the type of simulations used successfully in previous USAF research efforts, the laboratory's scenarios placed eight 3-person teams in a very dynamic and complex environment based on a fictitious regional conflict. To enhance the realism and salience of the scenarios for the crews, laboratory

researchers and technicians conducted reviews of regional events reported in the open press to design the overall scenarios. Researchers provided the WDs with a familiarization package summarizing the current state of "world events," as well as estimates of enemy and friendly combat resources (i.e., order of battle), established rules of engagement, deployment orders (i.e., Air Tasking Order), and mission objectives for each WD position. The crews were briefed on the general contents of the package and their importance to performing well in the simulation, but it was left to the subjects to decide if they would actually review the materials. This was done to stimulate team interaction outside of the laboratory, and to provide the subjects an additional opportunity to engage in team-building activities. Neither time nor resources allowed a more directive approach to providing the participants with general background information and mission descriptions. Each team was brought into the facility separately, completed the six scenarios, and returned to base before the next team departed. A schedule was established to maximize the time the teams were available to researchers. Upon completion, each team was debriefed about their participation and the intent of the project. They were directed not to disclose or discuss any aspect of the project to their compatriots at Tinker AFB until all teams had completed all the simulations. Based on individual and organizational feedback and the varied performance levels exhibited by the teams, the researchers felt the participants did not influence the results by discussing the project with their friends.

4. MEASURES

4.1. *Cue Ambiguity Manipulation.*

Each decision event was initiated by a scripted-communication (verbally transmitted) cue from a confederate playing the role of a friendly pilot. By reviewing the audio recordings of the incoming and outgoing transmissions, SMEs were able to verify both transmission and receipt of the cueing information. To increase the degree of interdependency associated with the task, cues were intentionally presented to a WD other than the one (target) for which its information was most appropriate. For example, the subject acting in the strike position received a request for emergency refueling from a damaged fighter aircraft. For the team to effectively deal with the situation, Strike would have to pass the information on to the HVA WD. At a minimum level of effective response, each decision event required the receiving WD to recognize and pass the cued information to the correct targeted position. Once the information was in the hands of the targeted position, a decision could be made and executed. While the target WD could handle some decisions, it was still important that individual team members remained aware of the consequences of these decisions for the other positions. Many of these decisions could have important consequences later in the scenario as resources and decision options became more restricted. Not all decision events required any actions other than recognizing and transferring the cueing information. Cue ambiguity levels were manipulated by using predetermined key words, reflecting current AWACS training procedures and SME experience, to indicate various levels of threat certainty (see Table 1).

Table 1: Cue Ambiguity Manipulation

	LOW	MEDIUM	HIGH	CERTAIN
Voice Message	"ESTIMATE"	"POSSIBLE"	"PROBABLE"	(as stated)
Certainty Level	25-50%	50-75%	75-90%	90-100%

4.2. *Environmental Complexity.*

To maintain a high level of task complexity, the scenarios were designed to present the teams with unexpected and multiple inputs. For example, some decision events included conflicting radar transmissions, which exploited the teams' heavy reliance on these transmissions for verifying cues and complicated the teams' collective picture of the situation. In light of

spurious electronic transmissions and mistaken labeling in the field, it was important to see if the team's coordination and situational awareness were sufficient to accurately discern the identity of the contact and correct any inaccurate electronic designations. By employing hostile radar jamming, researchers could also deny important information to the WDs until the team took appropriate action against the jammers. As in actual combat situations, such jamming is expected and is used to hide critical information indicating the full extent and size of the attacking wave from an adversary. The relatively high complexity level was expected to increase the teams' activity levels by forcing them to evaluate and prioritize multiple task demands, engage in collective problem solving, and carefully allocate limited resources to meet the situational demands. Research designers and technicians also believed this procedure enhanced the realism and operational relevance of the simulation scenarios.

4.3. *Process.*

Team process data were collected by SMEs observing the video- and audio-taped performance of each team. Because AWACS crews implement most of their decisions verbally, operational instructors and evaluators have come to heavily rely on the audio transmissions and verbal exchanges of crews to assess the reasons for their actions and decisions. In the case of the data supporting this proposal, audio-taped recordings played an especially important part in the SMEs' evaluations. Much like video recordings used in other research examining team process (Ref 11; Ref 31; Ref 21; Ref 32), these audio tapes provided SMEs with a rich accounting of the team's behaviors and interactions. Applying the same rating scales as used in previous AWACS research at the AESOP facility, SMEs encoded their assessments of each team's ability to detect and recognize decision cues embedded in the simulation (i.e., situational awareness). They also evaluated how well team members transmitted important and decision relevant information among themselves and used the information to deconflict task responsibilities (i.e., communication/coordination). Finally, SMEs assessed the quality and accuracy of the team's decision making process, when compared with previously determined decision options. These assessments were converted to z-scores and aggregated to the wave level of analysis to permit comparisons across teams and raters. We selected the wave level of aggregation as our focal point because decision events, within waves, were somewhat intractable. The wave level of analysis yielded more reliable and comparable effectiveness indices.

4.4. *Effectiveness.*

Using a complex scoring algorithm developed specifically for the AESOP simulator, SMEs recorded complete point totals for each wave and scenario. The formulas used to calculate these scores are based on an in-depth review of AWACS systems data, USAF doctrine, operational policies, and current training and evaluation standards and priorities. These formulas assigned weighted values for each resource and outcome involved in the simulation. Such decision outcomes as kill ratios (i.e., the ratio of enemy destroyed to friendly aircraft lost), number of successful air refuelings, number of assets lost to friendly fire (fratricide), and the degree of penetration into friendly airspace achieved by hostile aircraft were significant indices of team performance that the computer program/algorithm tracked. Based on previous research and SME feedback, AFRL-ARS researchers have found that these measures of team performance reliably reflect the complexity of the task environment and those actions directly under the team's control (Ref 33).

Difference scores were computed for each team. These scores reflected the increase or decrease

in total score due to a team's actions during any particular wave. Because decision events within each wave were identical for each team, these scores provided an objective outcome measure that could be used to compare teams at the wave level of analysis. Thus, as suggested by Tesluk et al. (Ref 32), observational and archival measurements were used to assess the data's fit to our hypothesized model and compare overall team performance.

5. RESULTS

Repeated measures multiple regression analyses were used to examine the significance of the relationships depicted in the hypothesized model (Ref 34; Ref 35; Ref 36). The results of the hypothesis model were mixed yielding some support for the anticipated relationships (see Figure 3).

Specifically, decision accuracy did exhibit a positive relationship with team performance. The relationship between communication/coordination and decision accuracy was also supported by the data. Although the correlation between situational awareness and decision accuracy was positive and significant, when considered together with communication/coordination, it did not evidence any unique influence (although it approached significance). The lack of significance may also be attributable, in part, to the somewhat low power of these analyses. Unexpectedly, neither environmental complexity or cue ambiguity had any significant linear effects on any team process. However, and more importantly, they did evidence a multiplicative effect as related to situational awareness. Specifically, as illustrated in Figure 4 (following page), the detrimental effects of environmental complexity were most pronounced when they were combined with ambiguous cues. In fact, environmental complexity did not have a negative effect on situational awareness if decision cues were clear (low ambiguity). In other words, operating in a complex environment or having to decipher ambiguous cues did not degrade situational awareness in themselves. If they occurred together, then situational awareness suffered greatly.

Figure 3: Team Performance Model

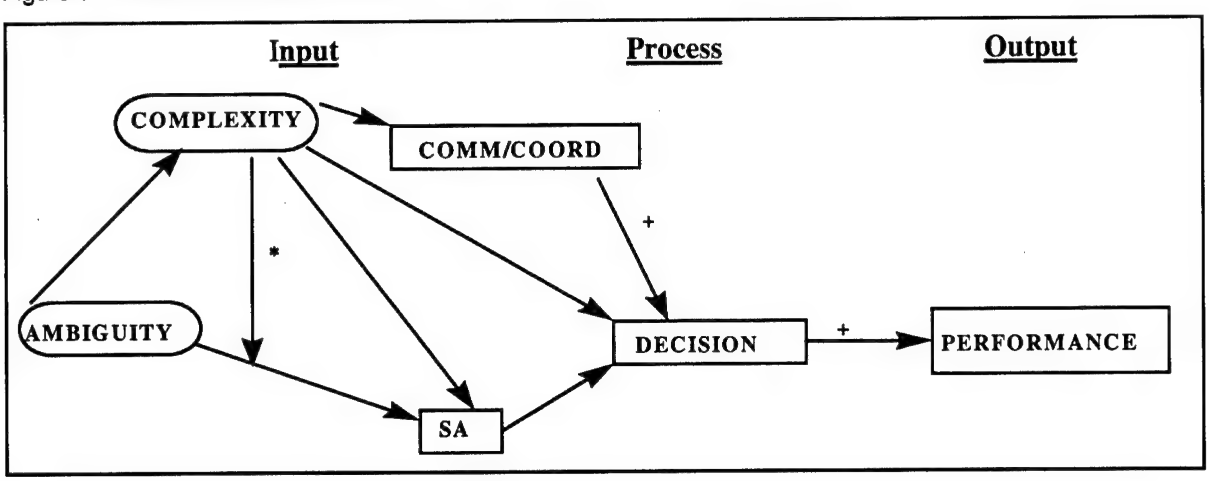
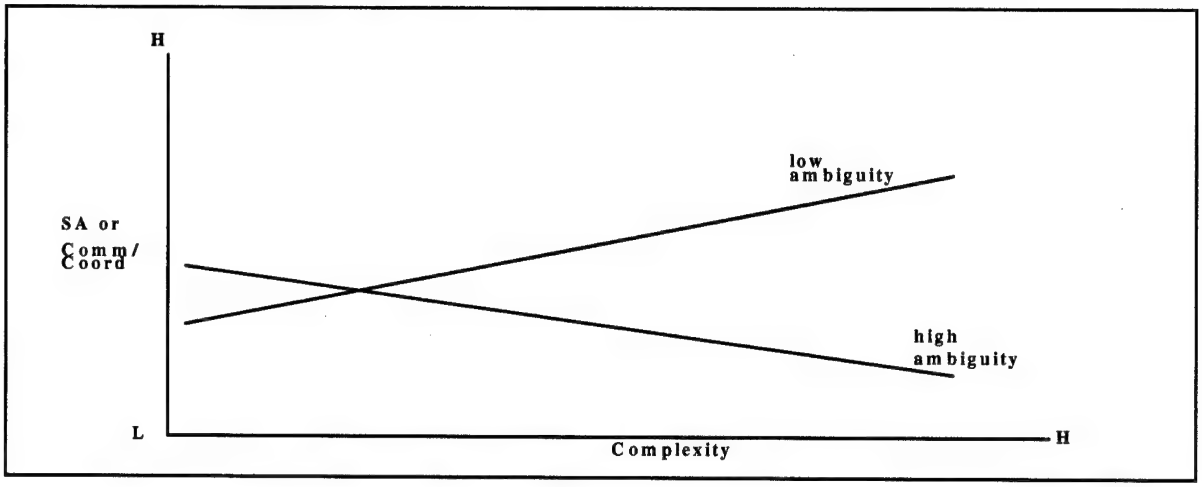


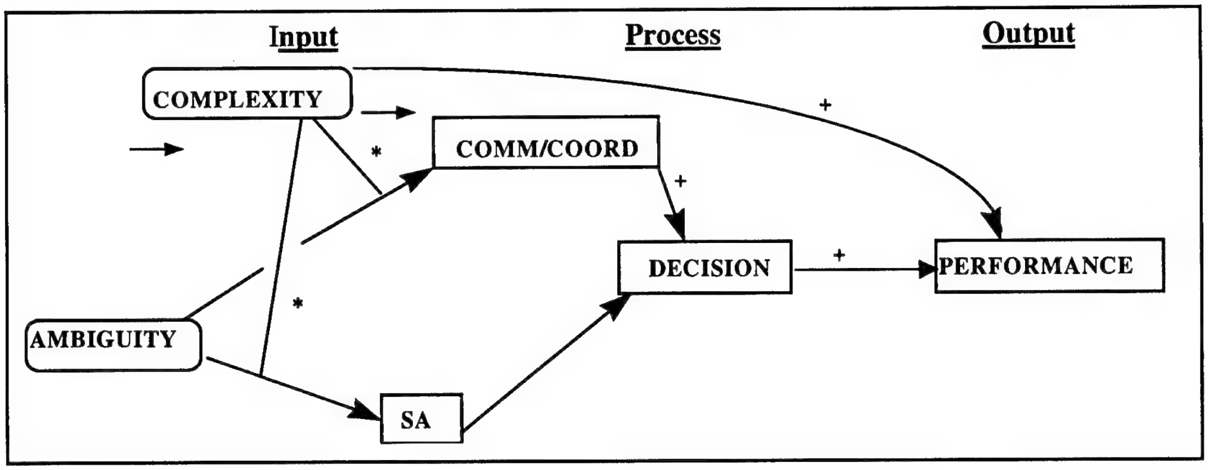
Figure 4: Cue and Complexity Effects



As noted earlier, environmental complexity did not evidence the significant linear effects on communication/coordination that were anticipated. However, in an exploratory analysis, we found that complexity interacted with cue ambiguity as related to communication/ coordination, just as it did with situational awareness. As complexity increased, the effect of cue ambiguity on each of the process variables became more pronounced and significant. The ability to detect and recognize increasingly ambiguous cues was degraded by increasingly complex situations. However, as could be expected by previous research (Ref 27; Ref 9), low cue ambiguity seemed unaffected by increased complexity levels. Clear cue presentations enhanced both situation awareness and team coordination/communication regardless of the level of complexity. In other words, complexity in and of itself did not degrade communication/coordination, but if combined with ambiguous cues, it did have deleterious effects.

Two other analytic points are worth noting. First, our hypothesized model had included a path from cue ambiguity to environmental complexity as a potential control for the modeling of environmental influences. As it turns out, these factors were found to be independent aspects of the performing environment. Stated differently, the quality of the environmental cues were not seen as contributing to the complexity of the environment, suggesting that the latter was considered more a function of the number of entities presented. Unexpectedly, environmental complexity did have a significant positive effect on performance. In retrospect, this makes sense as higher performances are attributable largely to the number of enemy forces neutralized and friendly forces retained, which becomes easier to achieve when more assets are present. As the number of entities involved in the simulation increase, raters would also observe greater levels of activity. Accordingly, a final revised model is presented in Figure 5 (following page) that includes the "control" effect of modeling a direct effect of environmental complexity on performance, the communication/coordination → decision accuracy → team performance sequence, and the interactive effects of environmental complexity and cue ambiguity on both communication/coordination and situational awareness. Finally, although not statistically significant, we maintained the situational awareness → decision accuracy linkage because of its theoretical importance and the fact that it neared statistical significance.

Figure 5: Team Performance Model



6. DISCUSSION

In many ways, the interactions observed in this data analysis confirm existing perceptions about the negative effects of ambiguity and environmental complexity on team decision processes. It was fairly unexpected that the interactions between cue ambiguity and environmental, rather than their main effects, provided the only significant influences on the team process measures used in this research effort. It was also interesting that low levels of ambiguity would be relatively unaffected, rather than being overwhelmed, by higher levels of complexity as indicated by crew activity levels. We believe that this result could be at least partially attributed to the fact that the teams were comprised of well-trained and experienced weapons directors. Due to their training and experience, team members were able to detect and recognize somewhat (low) ambiguous decision cues even when immersed in a highly complex environment.

The fact that decision cues followed team member expectations and established training standards could have also contributed to the effect noted in this analysis. Although verbal cues indicating the certainty an observer has of their perceptions is important in training and standardizing C^3 procedures, it may be the very nature of the cue that determines its true level of ambiguity. According to this reasoning, it might be better to manipulate cueing events in terms of their compliance with team expectations established during operational training. Unfortunately, the data supporting this research were not adequate to isolate any differences between various verbal and visual cues that do not directly conform to existing team expectations developed during training. Further research is required to demonstrate how varying levels and types of ambiguity and complexity impact team process.

As discussed earlier, situational awareness seemed to have no direct effect on decision quality despite a large amount of research supporting such a relationship (Ref 18; Ref 37; Ref 38). We concluded that this may be due to the close conceptual relationship between the measures used in this research for communication/coordination and situational awareness. The conceptual similarity between detecting a cue and informing others about it may be too high for our model, based on SME ratings, to isolate the amount of variance due to each variable. This could be causing the two variables to vie for the same portion of variance within the model. Research should employ alternate indices for these team processes, such as those used by Endsley (Ref 18), Goodwin, (Ref 38), and Hollenbeck et al (Ref 39), to

further examine the relationship between these process variables.

Unexpectedly, our exploratory analysis also indicated that environmental complexity had a significant main effect on teams' performance outcomes. We believe that this may be reflecting a control relationship, since the mechanisms governing activity levels were the same inputs for the scoring algorithms. For example, subjective SME ratings of team activity levels, used as our indicator of complexity levels, were based on their observations of all the team member's individual and collective actions. Most of these same actions, such as conducting switch actions to identify and track aircraft, were also assigned a weighted value by the computerized scoring algorithm. These values contributed to the team's overall performance scores for each wave and scenario and would be expected to increase as the number of individual tasks increased during the scenario. The relationship noted in our analysis could be due more to the nature of the computation formulas used to calculate performance scores than any direct linear relationship between complexity and performance. The increased number of aircraft alone may provide the opportunity for increased performance scores rather than differences in the teams' decision processes (Ref 40). Further examination and testing of the individual components and formulas comprising the scoring algorithm is required to adequately explain the origin of this unexpected finding.

This research has some interesting implications for operational training. Based on our findings, we suggest that simulation and training designers need to carefully consider the lack of any significant main effect for cue ambiguity and environmental complexity on team processes. We contend that any training aimed at improving team process, as well as future research in this area, must consider the interaction effects observed here rather than relying on isolated measures of ambiguity or complexity alone. This is especially true for training concentrating on improving team performance at higher levels of complexity.

The results from this research also seem to question the utility of assigning point values to every action a team member takes in a simulation and uses the resulting composite score as an overall indices of performance. While there is potentially much to be gained by isolating and operationally weighting each individual action in a simulator, designers should carefully consider what actions are actually relevant to the desired performance outcome. Trained and experienced SMEs may implicitly capture more of the relevant aspects of team performance than is possible with the strictly objective tracking measures employed in this research. Numbers alone may provide an incomplete picture of team performance and interaction. Future research could provide additional opportunities for comparing the effectiveness of subjective performance ratings and objective computer scores in assessing performance outcomes in both novice and experienced teams.

Based on our findings, designers of such complex scoring algorithms should also carefully consider how underlying scoring constructs and formulas replicate the measures employed to assess the impact of any predictor variables. If these measures are not sufficiently distinct from one another, then it may seriously undermine the conclusions drawn about what factors contributed to a team's performance. It may also be the case that general scoring algorithms, which weight every team member action, may not be desirable if one is attempting to understand what processes underlie the performance. These analyses may require a more finely collaborated or tailored outcome measure.

7. REFERENCES

1. "Conduct of the Persian Gulf War", Pursuant to Title V of the Persian Gulf Conflict Supplemental Authorization and Personnel Benefits Act of 1991, Public Law 102-25, April, 1992.
2. Owens, W.A., "The emerging system of systems", U.S. Naval Proceedings, May, 1995.
3. Owens, W.A., "A report on the JROC and the revolution in military affairs", Marine Corps Gazette, August, 1995.
4. Weaver, J.L., Bowers, C.A., Salas E., & Cannon-Bowers, J.A. "Networked simulations: New paradigms for team performance research", Behavior Research Methods, Instruments, & Computers, 27(1), 1995, pp. 12-24.
5. Brannick, M. T., Roach, R. M., & Salas, E. "Understanding team performance: A multimethod study", Human Performance, v6, 1993, pp. 287-308.
6. Morgan, B. B., Jr., Glickman, A. S., Woodward, E. A., Blaiwes, A. S., & Salas, E., "Measurement of team behaviors in a Navy environment", NTSC Technical Report, No. TR-86-014, Orlando, FL., Naval Training Systems Center, 1986.
7. Salas, E., Dickinson, T.L., Converse, S.A., & Tannenbaum, S.I., "Toward an understanding of team performance and training", In R. W. Swezey and E. Salas (eds.), Teams: Their Training and Performance, Norwood, NJ., Ablex, 1992, pp. 3-29.
8. Guzzo, R.A., & Dickson, M.W., "Teams in organizations: Recent research on performance and effectiveness", Annual Review of Psychology, v47, 1996, pp. 307-338.
9. Hackman, J.R., "Group influences on individuals in organizations", In M.D. Dunnette & L.M. Hough (Eds.), Handbook of Industrial and Organizational Psychology, 2nd Ed., v3, Chicago, IL., Rand-McNally, 1992, pp. 199-267.
10. Hackman, J.R., & Oldham, G.R., "Work Redesign", Reading, MA., Addison-Wesley, 1980.
11. Heffner, T.S. "The impact of task and team skills training on the relationship between mental models and team performance", Unpublished doctoral dissertation, The Pennsylvania State University, University Park, 1996.
12. Steiner, I. D. "Group processes and productivity", New York, Academic Press, 1972.
13. Mathieu, J. E., & Button, S. B. "An examination of the relative impact of normative information and self-efficacy on personal goals and performance over time", Journal of Applied Social Psychology, v22, 1992, pp. 1758-1775.
14. Tannenbaum, S.I., Salas, E., & Cannon-Bowers, J.A., "Promoting team effectiveness", In M.A. West (Ed.), Handbook of Work Group Psychology, John Wiley & Sons, Ltd., 1996, pp. 503-529.
15. Kozlowski, S.W., & Gully, S.M., "TEAMS/TANDEM: Examining skill acquisition, adaptability, and effectiveness", Presented at the Academy of Management Annual Meeting, Cincinnati, OH., 1996
16. Prince, C., & Salas, E., "Training and research for teamwork in the military aircrew", In Wiener, E.L., Kanki, B.G., & Helmreich, R.L. (Eds.), Cockpit Resource Management, San Fransico, CA., Academic Press, 1993, pp. 337-366.
17. Wellens, A. R., "Group situation awareness and distributed decision making from military to civilian applications", In N. J. Castellan, Jr. (ed.), Individual and group

decision making: Current issues, Hillsdale, NJ., Lawrence Erlbaum, 1993.

18. Endsley, M. R., "Towards a theory of situation awareness", Human Factors, v37, 1995, 32-64.

19. Stinnet, T. A., "Human factors in the super cockpit", In R. Jensen (Ed.), Aviation Psychology, Brookfield, VT., Gower, 1989.

20. Shrestha, L. B., Prince, C., Baker, D. P., & Salas, E., "Understanding situation awareness: Concepts, methods, and training", In W. B. Rouse (ed.), Human/technology Interaction in Complex Systems , v7, Greenwich, CT., JAI Press, 1995, pp. 45-83.

21. Mathieu, J.E., Goodwin, G.A., Heffner, T.S., & Born, D.H., "A longitudinal test of co-evolving team processes and environmental demands as related to team performance", Unpublished manuscript, Pennsylvania State University, 1996.

22. Gladstein, D., "Groups in context: A model of task group effectiveness", Administrative Science Quarterly, v29, 1984, 499-517.

23. Ancona, D.G., & Caldwell, D.F., "Beyond task and maintenance: Defining external functions in groups", Group and Organization Studies, Vol. 13, 1988, 468-494.

24. Ancona, D.G., & Caldwell, D.F., "Demography and design: Predictors of new product team performance", Organization Science, v3, 1992, 321-341.

25. Guzzo, R.A., & Shea, G.P., "Group performance and intergroup relations in organizations", In M.D. Dunnette & L.M. Hough (Eds.), Handbook of Industrial and Organizational Psychology, Chicago, IL., Rand-McNally, 1992, 269-313.

26. Saavedra, P., Earley, C., & Van Dyne, L., "Complex interdependence in task-performing groups", Journal of Applied Psychology, 78(1), 1993, pp. 61-72.

27. Eysenck, M.W., & Keane, M.T., "Cognitive Psychology: A Student's Handbook", 3rd Ed., East Sussex, UK., Psychology Press Publishers, 1995.

28. Elliott, L.R., Neville, K., & Dalrymple, M.A., "C^3Simulation, Training, and Research Systems (C^3STARS): Team Performance Under Conditions of Information Ambiguity—Constructs and Measures", Proceedings of the Annual Conference of the Human Factors Society, Philadelphia, PA., 1996.

29. Dalrymple, M., Personal communique, AESOP Facility, Brooks AFB, TX., Veridian, Sept. 23, 1997.

30. Fowlkes, J., Lane, N., Salas, E., Franz, T., & Oser, R., "Improving the measurement of team performance: The TARGETs methodology", Military Psychology, v6, 1994, 47-61.

31. Mathieu, J.E. & Goodwin, G.A., "A longitudinal examination of team efficacy - Performance regulatory mechanisms", Presented at the 12th annual conference of the Society for Industrial/Organizational Psychology, St. Louis, MO., 1997.

32. Tesluk, P., Mathieu, J.E., Zaccaro, S.J., & Marks, M., "Task and aggregation issues in the analysis and assessment of team performance", In Brannick, M.T., Salas, E., & Prince, C. (Eds.), Team Performance Assessment and Measurement: Theory, Methods, and Applications, Mahwah, NJ., Lawrence Erlbaum Associates, 1997, pp. 197-223.

33. Komaki, J.L., "Behind the scenes: Field testing a measure of effectiveness for theater teams", In Brannick, M.T., Salas, E., & Prince, C. (Eds.), Team Performance Assessment and Measurement: Theory, Methods, and Applications, Mahwah, NJ.,

Lawrence Erlbaum Associates, 1997, pp. 227-251.

34. Cohen, J., & Cohen, P., "Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences", 2nd Ed., Hillsdale, NJ., Lawrence Erlbaum Associates, 1983.

35. Colquitt, J.A., Hollenbeck, J.R., & Gully, S., "Repeated measures regression: Decomposing variance in cross-level research", Presented at the Annual Meeting of the Society of Industrial and Organizational Psychology, Dallas, TX., 1998.

36. Hollenbeck, J.R., Ilgen, D.R., & Sego, D.J., "Repeated measures regression and mediational tests: Enhancing the power of leadership research", Leadership Quarterly, v5, 1994, pp. 3-23.

37. Foushee, H.C., "Dyads and triads at 35, 000 feet: Factors affecting group process and aircrew performance", In H.W. Orlady & H.C. Foushee (Eds.), Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop held in San Francisco, California, May 6-8 1986, NASA Conference Publication 2455, 1987, pp. 265-279.

38. Goodwin, G.F., "Situational awareness: The translation of mental models into team performance", Unpublished master's thesis, The Pennsylvania State University, Pennsylvania, 1997.

39. Hollenbeck, J.R., Major, D.A., Sego, D.J., Hedlund, J., Ilgen, D.R., & Phillips, J., "Team decision-making accuracy under difficult conditions: Construct validation of potential manipulations using the TIDE^2 simulation", In Brannick, M.T., Salas, E., & Prince, C. (Eds.), Team Performance Assessment and Measurement: Theory, Methods, and Applications, Mahwah, NJ., Lawrence Erlbaum Associates, 1997, pp. 111-136.

40. Elliott, L.R., Schiflett, S.G., & Dalrymple, M.A., "Modeling the decisionmaking process and performance of Airborne Warning and Control System (AWACS) weapons directors", Presented at the 1998 NATO Symposium on Collaborative Crew Performance in Complex Operational Systems, Edinburgh, UK., 1998.

Distributed Cooperative Planning

Patricia M. Jones
University of Illinois at Urbana-Champaign
Dept. of Mechanical & Industrial Engineering
1206 W. Green Street, Urbana IL 61801 USA

p-jones5@uiuc.edu
http://tortie.me.uiuc.edu/Jones.html

SUMMARY

This paper defines distributed cooperative planning and provides an overview of some of the broad issues in computer-supported cooperative work. These issues include sharing an information space, articulation work, and social presence. Next, the relationship between planning and execution (operations) is discussed. Finally, methods for the study of distributed collaborative planning are described, and two current studies of distributed collaborative planning are highlighted.

1. INTRODUCTION

Distributed cooperative planning is a collaborative human activity that is distributed over time and space. In particular, cooperative planning means that the participants share the joint goal to construct a plan (or parts thereof) and, frequently, to track its execution when it becomes operational. Planning itself is defined as the construction of an abstract representation of activities, resources, and actors, to be coordinated and deployed over space and time.

The goals of the projects described in this paper are twofold: (1) the analysis and modeling of the sociotechnical processes involved in performing distributed cooperative planning and (2) the development and evaluation of prototype collaborative tools to support planning in situ.

This paper is organized as follows. Section 2 describes generic issues in the study of cooperative work systems. Section 3 focuses on planning and the transition from planning to operations. Section 4 discusses methodological issues, and Section 5 summarizes two ongoing research projects on collaborative planning.

2. THREE ASPECTS OF COOPERATIVE WORK

Per Schmidt and Bannon (1992) and as extended by Jones (1996), three aspects of cooperative work are sharing an information space, articulation work, and presence. That is, the analysis of a cooperative work system, and the design of collaboration technologies, must take into account three broad classes of issues: how to share information, how to coordinate activity, and how to manage presence of actors in a collaborative space.

2.1. Sharing an Information Space

Information sharing is a ubiquitous aspect of cooperative work systems. Key issues to be addressed include the tension between local, private representations and global, public ones. Some CSCW systems assume that everybody always needs to see the same thing (e.g., the "relevant common picture" of the Army's digital battlefield), while other systems

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

support local tailoring of work which then gets published to a group later. Many policy and architectural issues thus revolve around information sharing: what are the kinds of information in terms of form and content? which participants are allowed to see what? who has permission to read, write, or modify information or objects? how do we know what is relevant? how are heterogeneous views coordinated?
Thus, the links between roles of participants and their associated activities and authority with respect to the creation, modification, deletion, and sharing of information is critical. Issues of permissions, authority, and access control must be considered in light of the current work context.

CSCW architectures vary in their degree of centralization, which has implications for policies of sharing information. A centralized architecture maintains an authoritative version of information with locking or merging mechanisms. A decentralized architecture allows multiple competing versions to exist, and may support merging as well.

While it is important to consider technological issues in information sharing, the flexible and negotiated character of the social practices of information sharing are also crucial. The analytic concept of boundary objects has been used to analyze cooperative work and to design shared information spaces (Star, 1989; Star and Griesemer, 1989; Chin, 1997). Rather than assuming participants shared identical representations or constructed identical meanings, Star views partial shared meanings as an effective middle ground, where 'boundary objects' such as repositories, labels, and ideal types maintain enough identity to be coherent across different work contexts, but are plastic enough to adapt to local needs.

2.2. Articulation Work

A second aspect of cooperative work is articulation work: the invisible meshing and coordination of activity (Schmidt and Bannon, 1992). Two aspects of the coordination of activity are task allocation and redirection of attention. The definition of tasks, their assignment to participants, and the tracking of their status constitute task allocation. Support for ostentive behavior (pointing) and other nonverbal cues such as mutual gaze awareness are constituents of redirection of attention.

2.3. Social Presence

A third key facet and a second type of 'overhead' for cooperative work is the social work of maintaining a presence or identity. In some systems, such as traditional databases, there may be no explicit representation of identity at all; collaboration may occur simply through changes to data in persistent store. However, most CSCW systems allow some form of identification. Identity may be unknown or known; participants may be anonymous or be known as a personal name, title, rank, or member of an organization. Whether or not identity is public knowledge, participants still vary by perspective, goals, agenda, and authority.

Goffman (1959) describes in detail the kind of 'overhead' involved in social interaction. He views social interaction as impression management in which people design interactions to give and give off intended impressions. His analysis is based on the metaphor of the theater, in which a performer, or team of performers, performs front-stage behavior for an audience, designing manner, appearance and setting, and constructs, plans, and critiques this performance backstage.

Thus, in addition to technical issues of authentication, permissions and access control, and the design of avatars, are the social practices that arise around, through, and with collaboration technologies.

3. From Planning to Operations

Collaborative planning, as described in Section 1, is the collaborative construction of an abstract entity, the plan, that will be used to coordinate operations at some later time. The two projects described in Section 5 focus on supporting this collaborative planning type of process. However, the move from planning to execution, and associated issues in dynamic replanning, is particularly interesting. What is the status of the plan in this dynamic complex context?

Jones (1995) provides a discussion of how planning differs from operations. Planning can be seen as an example of a self-paced, creative activity in which 'undoing' is possible. In contrast, real-time operations are event-driven, dynamic, tactical, and not always 'undoable'.

Three ways that plans relate to operations are: plans as resources for action, planning as an opportunity for training, and planning as an opportunity to get to know team members and establish trust, expectations, social norms, working procedures, and the like.

The theory of situated actions views plans as resources for action, rather than the sole generating mechanism for behavior (Suchman, 1987). In this sense, the plan is consulted during execution but does not necessarily provide strong constraints on action.

A related view is that planning is useful as a training regime rather than as creation of an artifact that will be actually used during execution (Dunmire, 1997). Many plans are generated and never or rarely consulted during execution, but rather than saying this shows a failure in planning, we might also think through the kinds of knowledge and skills that people are better able to use in context given that they did participate in a planning process.

Thirdly, collaborative planning provides a joint activity in which participants get to know each other. This social justification for collaborative planning is significant in high-risk situations in which trust and appropriate knowledge of team members' expertise is a critical factor in effectively working together. The plan can further be seen as partial background construction of 'team situation awareness'.

4. Methods

The methods by which one studies collaborative planning come from a mix of sociology, anthropology, human factors engineering, and psychology. These include ethnographic techniques for naturalistic observation and analysis of archival material; interviews and questionnaires; and iterative prototyping of designs in the context of realistic use cases or scenarios.

The conceptual objects of study are participants, activity, information, and artifacts (Jones, 1995; Jones and Jasek, 1997). Thus, data are organized around building models that explicate who engages in what activity(ies), what information is used or transformed in those processes, and what artifacts are used or produced to guide activity. We view information as read-only and temporary and artifacts as read-write and durable.

5. Current Studies

Two current collaborative planning studies are in the military domain: CoRAVEN is for collaborative intelligence planning and analysis, and MUDSPOT/MAST is for collaborative planning for operations other than war, in particular, humanitarian assistance and disaster relief.

The CoRAVEN project is the study of, and development of a software prototype for, collaborative intelligence analysis in the United States Army. The general process of intelligence analysis is to

generate requirements for data gathering, plan the data collection effort, and then monitor messages received from the collection assets. The process starts with information requests from the commander, which are the questions that the intelligence staff will try to answer based on the data collected.

The goal of the CoRAVEN project is to explore technology to support this process, and we are particularly interested in using Bayesian belief networks as a structure of inference, collaborative graphical interfaces to support collaborative planning, and auditory cues that also support situation awareness. Thus, in CoRAVEN, we use Bayesian belief networks as the inference mechanism to reason about how messages relate to information requests. Our user interface is collaborative and offers a number of resources, including visualization and sonification of the belief networks. In terms of collaboration support, we will rely on active database technology (specifically, the POET commercial object-oriented database) to implement a flexible collaboration policy, so that users can change their level of collaboration at run-time.

The MAST/MUDSPOT project is part of the C2MUVE project (Duffy, 1998) in which we are supporting the Operations Planning Team in doing crisis action planning for humanitarian assistance and disaster relief applications. The Mission Analysis Support Tool (MAST) is the front end to the MUDSPOT architecture that is based on a freeware Java MUD (Multi-User Dungeon) and our SPOT project/activity management prototype (Jacobs et al., 1997). We have a rich representation of objects for collaborative planning, such as Activity, TimePoint, Project, and Participant, that is based on the Shared Planning and Activity Representation (SPAR), and we have implemented the SPAR specification in Java for our purposes. Our specific task in the context of the larger C2MUVE project is to design and implement simple workflow coordination mechanisms.

6. References

1. Chin, G. (1997). Management of boundary objects in a shared information space for a public works organization. Ph.D. dissertation, Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign.

2. Duffy, L. (1998). Command and control multi-user virtual environment: Virtual support for operations other than war. Presentation at NATO HFM Symposium on Collaborative Crew Performance in Complex Operational Systems, Edinburgh, April 1998.

3. Dunmire, C. (1997). Personal communication.

4. Goffman, E. (1959). *The presentation of self in everyday life.*

5. Jacobs, J. L., Dorneich, M. C., Jones, P. M., O'Keefe, B. J. and Contractor, N. (1997). SPOT: Using collaborative technologies for developing collaborative technologies. *Proceedings of the 1997 IEEE International Conference on Systems, Man, and Cybernetics, Volume 1*, pp. 690-695. Orlando FL, October 1997.

5. Jones, P. M. (1995). Cooperative work in mission operations: Analysis and implications for computer support. *Computer-Supported Cooperative Work: An International Journal, 3,* 1, 103-145.

6. Jones, P. M. and Jasek, C. A. (1997). Intelligent support for activity management (ISAM): An architecture to support distributed supervisory control. *IEEE Transactions on Systems, Man, and Cybernetics*, Special issue on Human Interaction in Complex Systems, Vol. 27, No. 3, May 1997, 274-288.

7. Schmidt, K. and Bannon, L. (1992). Taking CSCW seriously: Supporting articulation work. *Computer Supported Cooperative Work, 1*, 7-40.

8. Star, S. L. (1989). The structure of ill-structured solutions: Boundary objects and heterogeneous distributed problem solving. pp. 37-54 in *Distributed Artificial Intelligence, Volume II* (eds. L. Gasser and M. Huhns). Morgan Kaufmann.

9. Star, S. L. and Griesemer, J. (1989). Institutional ecology, 'translations', and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*, 387-420.

10. Suchman, L. (1987). *Plans and situated actions: The problem of human-machine communication.* Cambridge University Press.

7. Acknowledgments

Many of the ideas in this paper grew out of collaboration with Dr. LorRaine Duffy while I was a Summer Faculty Fellow at NRaD in the summer of 1996. The MUDSPOT project is supported by NRaD Contract No. N66001-97-M-2077. The CoRAVEN project is supported by the Army Research Laboratory (ARL) and Army Intelligence and Security Command (INSCOM) via the ARL Federated Laboratory on Advanced Displays and Interactive Displays hosted at the University of Illinois, Urbana-Champaign, Cooperative Agreement DAAL01-96-2-0003. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied of the Army Research Laboratory or the U.S. Government.

Command and Control Multi-User Virtual Environment:

Virtual Support for Operations Other than War

LorRaine Duffy, Ph.D.
SPAWARSYSCEN D4123
53560 Hull Street
San Diego, CA 92152-5001, USA
lduffy@nosc.mil

Technical Objective

Provide a smart ,virtual environment (for Operations Other Than War) that integrates customized, automated work mechanisms with distributed, human, expert activities in the conduct of distributed, collaborative planning and execution monitoring.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

Operational Setting

C3 Collaborative Workspace | Virtual Environment Thrust Area

- CINC (Commander in Chief) level planning
  - » (Strategic, not tactical)
- First order situation assessment, planning, and response for theater level action
- Operations Planning Team (8-12 core members, 20-100 support personnel)
- Typically has few hours to few days to formulate first level plan and choose the Commander, Joint Tactical Force (CJTF)

Operational Payoff

C3 Collaborative Workspace | Virtual Environment Thrust Area

- Electronically support multiple planners in a distributed, multi-expert, collaborative task (e.g., producing Theater-level plan)
- Increase *consistent* situation understanding among distributed planners with the incorporation of groupware functions
- Shorten the current CINCPAC Operational Planning Team cycle (production of their planning products) from 24-96 hours to 4-24 hours

SPAWAR
Systems Center San Diego

4 Task Objectives

C3 Collaborative Workspace | *Virtual Environment Thrust Area*

Focus year

1. **MUD/MOO Architecture** FY96

   Develop multi-person cyberspace environment that exploits and builds from the current multi-user domain (MUD)/object oriented (MOO) academic environment;

   LambdaMoo w/ Java Applets (Netscape 3.0/4.04)

2. **Groupware Functionality** FY97

   Add commercial groupware functionality to support remote, distributed group decision making and planning activities

   (Facilitate.com, Groupsystems, Habanero,IVOX; currently assessing several new products)

SPAWAR
Systems Center San Diego

4 Task Objectives

C3 Collaborative Workspace | *Virtual Environment Thrust Area*

3. **Graphical Extensions** FY98

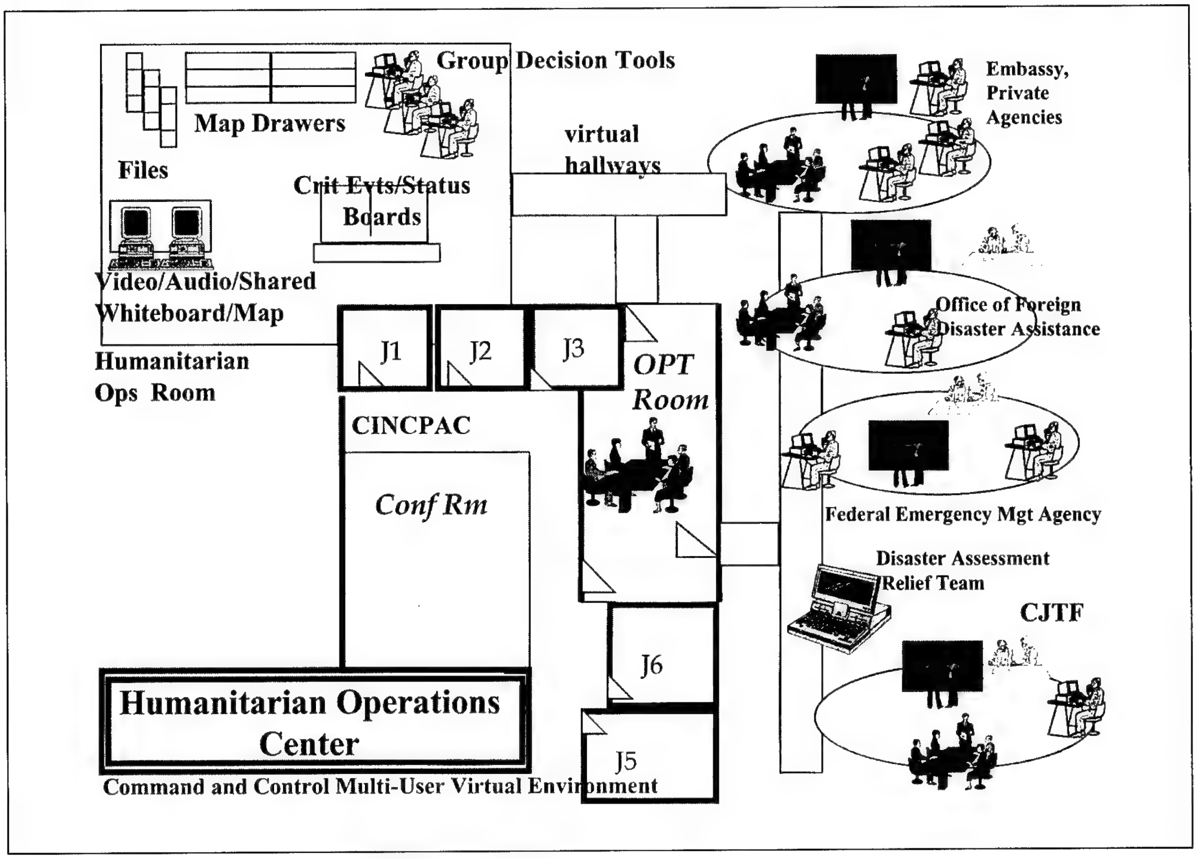
   Develop graphical extensions to MUDs/MOO (using VRML2.0-virtual reality modeling language)

   » **Additional HCI (speech and gesture recognition) technologies**

   » **context representation (increasing metaphors for interaction)**

4. **Intelligent Agents** FY99

   Add intelligent agent architectures that support workgroup functions, not just single user functions (e.g., navigation (guide), facilitation, transcription support; workflow support)

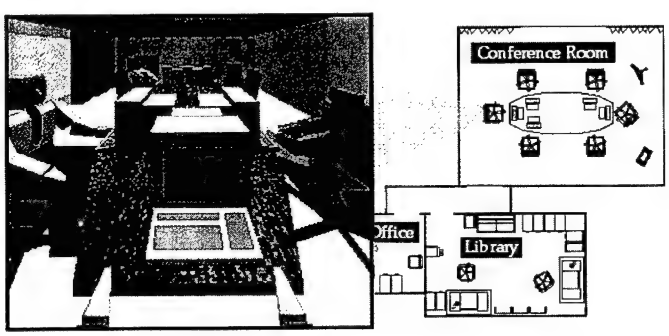


Holistic Virtual Environment

C3 Collaborative Workspace | *Virtual Environment Thrust Area*

Systems Center San Diego

Prototype Graphical Display



- 3-D, Navigable scene (room or function metaphor)
- Avatars - representations of other users
- Objects in rooms - table, computers, walls, groupware
- Overview maps to provide "location" context
- Communication/collaboration support mechanisms

Supported Human Activities

C3 Collaborative Workspace | Virtual Environment Thrust Area

- Communication
  - » Variety of channels/venue
- Collaboration
  - » Problem solving, team decision making
- Coordination
  - » Information exchange, Notification of work progress/status (workflow)
- Cooperation
  - » Acknowledgement of support, follow-through, change, or "disagreement"

Features

C3 Collaborative Workspace | Virtual Environment Thrust Area

- User-controlled virtual environment
- Locating and Paging other users (online or not)
- Email Users
- User centered Navigation map
- User-built workspaces
- Users able to build links to doc/ppt/urls/objects in workspaces
- User controlled system admin functions
- Interactive, color coded, public-private text chat
- Automated audio services (within a platform)

Related Work

C3 Collaborative Workspace

Virtual Environment Thrust Area

- MITRE's Collaborative Virtual Workspace/JFACC Collaboration Services (DARPA funded: Peter Spellman)
- U of Illinois' Orbit (previously wOrlds) (DARPA funded: Simon Kaplan)
- Several Commercial/Research Environments
  - » Alpha World to WorldsAway

Research Gaps

C3 Collaborative Workspace

Virtual Environment Thrust Area

We need further research on:

- **effective metaphors and concepts**
- **voice communications in Web based environment**
- **security issues**
- **little theory on effective visual environment for social behavior/cognitive load on strategic dec mkg tasks**
- **avatar research**
- **Virtual environment HCI has limited development**
- **real time interactive environments**
- **application on Demand (Wrapping Technology)**
- **navigation/browsing theory**

Summary

C3 Collaborative Workspace

Virtual Environment Thrust Area

Science &Tech Contributions

- Building the testable virtual environment for future theoretical work in this area
- Catalog
- Integration of Raw Technologies:
  - » MUDs/MOOs to support 100's of participants
  - » Graphical Extensions to MOOs
  - » Real time CORBA (Common object request broker architecture)
  - » Intelligent Agents for groupwork
- Incorporation of commercial Groupware Functionalities (work group support mechanisms)
- Continuing Integration of "Other" New Technologies
  - » speech, gesture, and "emote" recognition (avatars)
  - » wrappers for multi-platform capability

THE ROLE THAT COGNITIVE ABILITY PLAYS IN CRM

Carl C. Hoffmann, Ph.D.
Kathleen P. Hoffmann
Hoffmann Research Associates
111 Providence Road
Chapel Hill, NC, 27514, USA
Gary G. Kay, Ph.D.
Georgetown Medical Center
Washington, DC, 20007, USA

SUMMARY

Crew Resource Management (CRM) is now an important component of most airline training. It is clear from the analysis of accidents that a dysfunctional team in the cockpit will cause accidents.

Research in this area had focused on the degree to which training can facilitate team work in the cockpit and the degree to which people may have the personality structure, instrumentality and expressiveness that would either promote or interfere with their ability to work as a team member. The research presented in this paper examines four potential components of CRM. These components are aviation knowledge, general intelligence, personality, and cognitive skills. In order to communicate effectively and be a part of the team, the crew member must: 1) know the subject matter that is the basis of work that the team must execute; 2) have the intelligence to understand dynamic situations, formulate and communicate an idea; 3) be willing to engage in an exchange of ideas; and 4) have the processing capability to do multiple tasks simultaneously and not be overwhelmed by them.

Our paper will present the results of a study of 115 first officers for a major airline. Our dependent variables are scales derived from an in-flight observation instrument design to measure pilot performance in several areas – CRM among them. Our independent variables are an extensive job knowledge test developed by the airline, several standard intelligence tests, a personality inventory and CogScreen. Our findings indicate that cognitive ability is the major component of predicting good performance on our CRM scale.

INTRODUCTION

For 20 years, airlines have recognized that communication and coordination of activities in the cockpit is critical to crew performance. A study by Cooper, White & Lauber (1979) (p. 207) (using (Ruffell Smith (1979)—simulator B-747 study) set up a series of problems for a crew on a simulated flight. They found that most crew performance problems were related to breakdowns in crew coordination, not to a lack of technical knowledge and skill. "High error" crews experienced difficulties in the areas of communication, crew interaction and integration. Other performance deficiencies were associated with poor leadership and the failure of the flight crew to exchange information in a timely fashion.

Foushee and Manor (1981) (p. 209) analyzed the cockpit voice recordings from the Ruffell Smith (1979) simulation study. Overall, there was a tendency for crews who communicated less to perform less well, but the type or quality of communication played an even more pivotal role. There was a negative relationship between

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

crew member observations about flight status and error related to the operation of aircraft systems. In short, when more information was transferred about aspects of flight status, fewer errors occurred relative to such problems as mishandling of engines, hydraulic, and fuel systems, the misreading and missetting of instruments, the failure to use ice protections, and so forth.

Researchers have generally focused on personality factors as the cause of the breakdown in crew communication. As a result, interventions have addressed personality factors in attempts at remedying the problem. Robert Helmreich has pioneered this effort, focusing on exploring positive and negative aspects of two personality dimensions, *instrumentality* and *expressiveness*. (Helmreich and Spence, 1978; Spence, Helmreich, & Holahan, 1979). Data on these two dimensions is obtained using a self-report instrument called the EPAQ (Extended Personality Attributes Questionnaire):

1. **Instrumental Traits** relate to achievement and goal seeking (achievement motivation)

Instrumentality (I+): a cluster of positive attributes reflecting goal–orientation and independence (active, self-confident, can stand up to pressure)

Negative Instrumentality (I-): negative characteristics reflecting arrogance, hostility, and interpersonal invulnerability (boastful, egotistical, dictatorial)

2. **Expressive Traits** relate to interpersonal behaviors, sensitivity, and orientation.

   Expressivity (E+): a cluster of positive attributes reflecting interpersonal warmth and sensitivity (gentle, kind, aware of feelings of others)

   Negative Communication (E-): self-subordinating, subservient, or unassertive characteristics (gullible, spineless, subordinates self to others)

   Verbal Aggressiveness (Eva-): verbal passive-aggressive characteristics (complaining, nagging, fussy)

Helmreich and his colleagues subsequently added three measures of achievement motivation. The entire inventory is known as the Personal Characteristics Inventory (PCI).

In 1986, Chidester, Helmreich and others collected data as part of a CRM training program from two samples of military pilots in order to explore and validate subgroups that could be distinguished from their configuration of personality characteristics. Three clusters were identified:

(1) **Positive Instrumental/ Interpersonal** cluster, characterized by elevated levels of both positive instrumental and expressive traits, and below average levels on Negative Instrumentality and Verbal Aggressiveness.

(2) **Negative Instrumental** cluster, characterized by elevated levels of positive and negative instrumental traits, high verbal aggressiveness, work master, and competitiveness, and low levels of positive expressivity. "In a sense, pilots fitting this cluster can be characterized as instrumental, but not at all expressive. Individuals whose traits resemble this pattern might be best described as rugged individualists rather than team players." (Chidester, 1986).

(3) **Low Motivation** (sometimes labeled "Negative Expressive") cluster, characterized by below-average scores on positive instrumental and expressive

> traits; they show some elevation in Verbal Aggressiveness and Negative Instrumentality. These pilots appear to be neither instrumental nor expressive, but show some elevation on Verbal Aggressiveness and Negative Instrumentality.

These clusters were replicated in a second sample of pilots. Chidester, et al., considered Cluster 1 optimal for the close interpersonal coordination required of crew members in multicrew aircraft (Chidester, et al., 1991). This same study reports also evidence that personality cluster membership may set some limits on the effectiveness of CRM training. Individuals in the Positive Instrumental cluster members appear to gain the most from training and Low Motivation cluster members gain the least.

In related research by Chidester et al. (1990), these clusters were examined in relation to pilot performance in full-mission simulations. Crews led by a Positive Instrumental captain were rated by observers as consistently effective and made the fewest errors during abnormal simulation conditions. Crews led by Low Motivation captains made significantly more errors and were rated as less effective. Crews led by Negative Instrumental captains performed more poorly initially, but by the last abnormal segment, these crews were performing as well as Positive Instrumental crews.

The willingness to work as a member of a team, the desire to listen and express oneself, the ability to subordinate one's ego while maintaining independent judgment, and the conscientiousness required to adhere to protocols of communication are clearly important determinants of cockpit performance. However, personality does not tell the whole story of effective CRM. Pilots must not only have the desire to communicate, they also have to have something to communicate, and be able to communicate while doing many other tasks.

Therefore, intelligence, experience, job knowledge, and processing capacity should all be related to CRM.

The relationship of general intelligence, "*g*", to pilot performance is as well documented as any relationship in aviation psychology. In a study of 1,400 undergraduate navigator trainees and 4,000 undergraduate pilot trainees, the best predictor of overall performance, including tests and performance in the aircraft, was g, followed by knowledge of aviation information and instrument comprehension (Olea and Ree, 1994, p. 848). In fact, when measuring individual performance, some prominent authors in aviation argue that other measures show only small and marginal incremental validity.

> Although the names and appearance of the tests used in pilot selection vary, most are measures of g. The incremental validity of specific knowledge (e.g., aviation information, comprehension of aircraft terms), psychomotor abilities and personality scores has been shown to be small but significant. (Ree and Carretta, 1995).

Experience also plays a role in CRM. In examining decision making, authors have found experience to be very predictive of performance (O'Hare and Wiggins, 1993). While much of this work has been done on individuals, other authors have pointed to the group dynamic in decision making in commercial aircraft. These authors recognize the interaction of individual abilities and styles and crew coordination. Mosier-O'Neill and others have recognized the interaction of individual abilities and styles and crew coordination—individual within a group (crew) within an organization (Mosier-O'Neill, 1989). There is a necessity among the crew for shared assessments, shared mental models, and

coordination of actions. As a consequence, one could expect experience in type of aircraft or training to affect how pilots approach CRM. Presumably, those pilots trained in equipment requiring crew coordination, i.e., military airlift command (MAC), may have more training and knowledge of CRM than individual fighter pilots.

Finally, effective performance as a pilot--and presumably in CRM--requires the ability to multitask, to do the job you are assigned while coordinating with others.

> A time-sharing ability refers to the ability to perform multiple tasks in combination. Assessing a time-sharing ability is important for selection and training in any given complex task situation. (Ackerman, Schneider, & Wickens, 1984, p. 71).

This belief is shared by other authors (Damos, 1993; and Imhoff and Levine, 1981, p. 74) and is supported by validation studies conducted with CogScreen-Aeromedical Edition (reported in Kay, 1995).

CogScreen is a computerized cognitive screening test originally designed for the U.S. Federal Aviation Administration as an instrument for evaluating pilots' neuro-cognitive fitness-to-fly. The selection of tests for CogScreen was based on existing task analyses of the cognitive and psychomotor demands of flying. Prior studies have demonstrated that CogScreen is not only sensitive to changes in brain functioning resulting from trauma, substance abuse or illness, but that it is also predictive of flight performance. The cockpit-observed performance deficits reported for pilots who had been referred for clinical CogScreen examinations have been shown to be reflected in their CogScreen performance (Kay, ASMA92). Studies of the relationship between CogScreen and flight performance have also looked at operational flight data. In one study, the frequency and severity of flight performance violations was measured by flight data recorders. The study (Yakimovich, ASMA95) demonstrated that variables from a small number of CogScreen tests (i.e., Dual Task Test, Shifting Attention Test, Divided Attention Test, Pathfinder, and Backward Digit Span) were able to account 30% to 45% of the variance in the performance violation index. Nearly all of these tests measure aspects of multitasking (i.e., divided attention, mental flexibility, planning, and sequencing).

THE STUDY

We have examined the extent to which personality dimensions, intelligence, experience, aviation knowledge, and cognitive capacity can predict performance on crew resource management tasks. Such a model could help in pilot selection, especially for commercial airlines that rely greatly on hiring pilots already trained in a number of different environments with varying backgrounds in intelligence, job knowledge, experience, training, and various levels of ability to multitask. Along these lines, we developed and conducted our study as part of a pilot selection process for a major United States airline.

In 1996 and 1997, this airline sought to validate its newly designed selection procedure. The major components of that selection procedure were measures of general intelligence, job knowledge, structured behavioral interviews, personality measures, and tests of cognitive ability. These measures were to be related to performance in training and performance on the job. The study included observations of performance in the cockpit as part of a line audit procedure, and a formal instrument was used as part of the line audit. The instrument was based upon a job analysis that established both the behaviors on the job and the knowledge, skills

and abilities required for those behaviors. Current pilots for the airline were to be tested on the proposed selection battery, and information was to be collected about their training and performance on the job. The components of the battery that were most predictive of the outcome measures were then be used as part of the new selection process. As part of the validation process, a contrasting group of low-hour, weekend general aviation pilots were used to help add variability on the proposed selection instruments.

POPULATION

The group that was studied and whose information is reported here were a set of randomly selected civilian airline pilots who had been hired by the airline between 1988 and 1991 and who, at the time of the study, had reached the position of First Officer. The vast majority of the pilots in this group were trained in either the Navy or the Air Force. They were further selected to be above the 50th percentile of intelligence for Air Force Captains, and in fact over 60% of this group was above the 80th percentile. Further, this group had been screened at the time of hire for both company and cockpit "fit." Pilots who were not interested in continuing to fly and who were viewed as having personalities that would conflict with the culture of the company or with crew resource management were selected out. Finally, these individuals were put through a psychological screen, and anyone displaying psychopathology or a highly defensive personality was also screened out. As a consequence, our study involved a very homogenous group of aviators. These people were all well trained, highly knowledgeable, very intelligent, almost all male, and almost all Caucasian.

VARIABLES

Outcome Measures

All training episodes from hire until June of 1997 were recorded for each of the 115 pilots in our study. Any evidence of problems in training was recorded. Difficulty in written or oral exams, extra time in the simulator or in training in general, and problems with proficiency checks were all noted. Our predictive variable was dichotomous:
0 = no problems; 1 = evidence of any problems. Fifteen percent of this group reported having some problem in training during their career at this airline.

This constrained variability led us to derive additional outcome measures of performance to validate the potential selection procedures. Therefore, a new line evaluation instrument was designed to satisfy the following criteria:

- Be more sensitive to differences in performance than current line check ratings, thus providing needed variance for validating selection devices;
- Be valid, that is, reflect behaviors that are part of a pilot's job at the airline;
- Be composed of items that are observable and thus can be reliably rated;
- Have scales that are easy to interpret;
- Be easy to use and capable of being filled out in real-time, on the flight deck.

The new evaluation tool was pre-tested and applied as part of a general line audit conducted at the airline. It was used to evaluate performance of First Officers comprising the validation sample, as well as a sample of Captains and First Officers covering the entire theater of operations at the airline.

The instrument was developed based on a job analysis performed for the selection process. Line Check Airmen participated in focus groups to revise task statements and provide ratings of the relevance of each of the task statements. The focus groups also worked to minimize the effect of the type of aircraft, time in aircraft, familiarity

with airport, and weather conditions. We further controlled these factors throughout the project through data collection and statistical controls, as well as rater training. Clearly defined standards were developed to rate each task and to aid inter-rater reliability. The instrument was initially tested with an industrial psychologist, the Chief of Flight Standards and the Chief of Line Check Airmen. Extensive training sessions were held with Line Check Airmen that concentrated on applying the rating scales uniformly and establishing procedures covering data collection and completion of the rating forms. Each of our 115 pilots was observed as pilot flying so as to compare performance of the same activities. This information was collected and analyzed not only for our 115 pilots, but also for the 1500 Captains and First Officers observed in line audit procedures. A factor analysis was done on the results and four factors emerged from the analysis (see Table 1). The items associated with CRM activities and their factor loadings are recorded in Appendix A. Predictor variables were derived from the following sources:

Tests of Ability and Aptitude
Knowledge-Based Tests
Personality
Cognitive Processing
Prior Experience

Tests of Ability & Aptitude

Numerical Ability – As measured by the Differential Aptitude Test

Verbal Ability – As measured by Differential Aptitude Test

Mechanical Ability – As measured by the Bennett Mechanical Aptitude Test

Spatial Ability – As measured by the Minnesota Paper Form Board Test

Test of Reasoning – As measured by the Watson-Glaser Test of Critical Thinking

Nonverbal Aptitude – As measured by Raven's Progressive Matrices

Personality – (see Costa and McCrae, 1992)

Neuroticism
Openness
Agreeableness
Conscientiousness
Extroversion

Knowledge-Based Tests

Knowledge of Aerodynamics
Knowledge of Engineering
Knowledge of Navigation
Knowledge of Meteorology
Knowledge of Aviation Physiology

Cognitive Processing – As measured by CogScreen subtests:

Math Accuracy
Visual Sequencing
Symbol Digit Coding
Matching to Sample
The Manikin Test
Divided Attention
Auditory Sequence Comparison
Pathfinder
Shifting Attention Test
Dual Tasking

Prior Aviation Experience

Fighter experience
Large jet air transport
Turbo prop

PROCEDURES

Tests were administered over a two-day period, with the knowledge-based test administered first. The other tests were given in six different sequences in order to prevent fatigue from systematically affecting performance. All tests were administered and scored by computer. The information was compiled and added to the line observation data and training information, as well as the pilot's prior experience before being hired by the airline. Only individuals with

complete information on all measures were used in the study, leaving 110 analyzable subjects.

RESULTS

The mean and standard deviations for the variables used in this analysis are shown in Table 2. A histogram showing the distribution of CRM scores is found in Figure 1. Clearly, this outcome variable has more variance than those involving simple pass/fail dichotomies.

In Table 3, we present the correlation coefficients for each independent variable with the CRM measures. Several findings are notable. Using a 0.10 significance level as a first indicator, we see that Divided Attention Sequence Comparison Speed is negatively related to CRM, while Divided Attention Sequence Comparison Throughput is positively correlated to CRM. The faster pilots make comparisons, the higher their CRM rating. Speed has the same relationship for the Manikin test, and Shifting Attention Test Arrow Color Condition. CRM is significantly correlated with two accuracy measures: Matching to Sample Accuracy and Shifting Attention Test Discovery Accuracy. On process measures, significant correlations were found for Shifting Attention Test Discovery Condition Rule Shifts Completed and Number of Failed Sets. The more successful pilots are at systematically and flexibly applying rules, the better they are at CRM. The Dual Tracking Test (Boundary Hits) also produces significant associations, but in unanticipated ways: the correlation is significant – but non-linear. The more tracking errors pilots made (within limits), the better they are at CRM. No intelligence tests or knowledge-based tests are significant. Furthermore, among the personality measures, only Agreeableness is significantly associated with CRM. Two experience variables are associated with CRM: Prior Fighter Experience is positive, while Jet Transport Experience is negative.

A regression model limited to CogScreen variables (with r values at the $p< 0.1$ level); past flight experience; Agreeableness; aviation knowledge (of aerodynamics, engineering and navigation); and the intelligence tests was used to find a predictive equation composed of variables that contributed significantly to CRM performance at the .05 level of significance. The final model arrived at through stepwise regression accounted for 35.3% of the variance in CRM. These results are shown in Table 4.

DISCUSSION

From the parameters of the regression equation, we see that there are many components that predict CRM. Some of the findings make sense intuitively, while others are surprising. As expectedly, Agreeableness is positively correlated with CRM. Agreeableness here is defined as trust, straightforwardness, altruism, compliance, modesty and tender-mindedness. Clearly, these personality dimensions are indicative of someone who can cooperate, although not necessarily subserviently. Knowledge of aerodynamics is positively correlated with CRM, which suggests that the more pilots know about the principles of flight, the better they can communicate. This knowledge helps facilitate communication in that it provides the substance of the communication and indicates a lesser degree of cognitive effort and analysis required of pilots who can quickly access and recall their knowledge.

Of the three measures from CogScreen, two have the expected relationship to CRM. Match to Sample Accuracy is positively correlated, while Divided Attention Sequence Comparison Speed is negative, meaning the more accurate pilots are, and the less time they spend reacting, the better their scores on CRM. This finding indicates that processing speed and accuracy facilitate CRM. The one inconsistent relationship from CogScreen is dual tasking: the more errors a pilot makes on this measure, the higher the CRM

score. This relationship could be explained by a particular test-taking strategy used by the better pilots. They are apparently willing to let some errors occur while they are optimizing both tasks on the test; in other words, focusing on perfect accuracy here may result in lowered performance in other areas.

We can see several other surprising results. Individuals with background in jet transport who have been instructors scored lower than those without these aspects of experience. In fact, this finding is somewhat ironic, since these individuals have been most likely exposed to crew situation during training. Also ironically, pilots with fighter experience and instructor experience had a positive correlation to CRM rating. (See Table 3.) This runs counter to the popular belief that fighter pilots are individualistic, egocentric, non-team players. It may be a result of the extensive screening for cockpit fit that went into selecting these individuals. Or, it could also be due to the military's practice of placing better students in the fighter pilot career path. Finally, there is a slight negative correlation of the Differential Aptitude Test (DAT) measure of overall aptitude and CRM. This relationship may be in part due to the poor upper-end differentiation of the DAT. Overall, this model explains 35% of the variance, which is quite high compared to similar studies. If the sum of squares is partitioned, the result is that 33% of the variance is explained by cognitive measures, 25% by experience, 16% by agreeableness, 13% by knowledge and 9% by intelligence.

From this study and this population, we see that it is important to look beyond personality to explain how pilots will perform at CRM. Clearly, knowledge, cognitive ability, prior experience and intelligence all play a role in CRM. The variables that play an important role in this study are also significant in other studies on how cognitive ability relates to performance, as we discussed in the Introduction. In this study, cognitive ability played the greatest role in predicting performance on CRM. Our findings indicate that not only must pilots have the kind of personality that can cooperate in the cockpit, but they must also have the knowledge and cognitive resources to perform well. Further, it is also clear that prior experience working with crews in large aircraft does not necessarily indicate good CRM performance.

It should be kept in mind that this group of pilots was a unique, relatively similar, group. Again, they had almost all been trained in the U.S. military; they all had to be at least at the median level for Air Force Captains; and they all had to pass rigorous psychological screening and personal interviews. But their homogeneity makes it even more surprising that we can see and explain as much variance as we can in such a group. The screen on intelligence may explain the negative relationship to the DAT found in the regression equation.

Cognitive ability, along with job knowledge, experience, personality, and general ability, play a role in other measures of performance. In Table 5, we present a grid that lists the components of our measures that predict positive performance in training and in two other measures of cockpit performance: procedural compliance and aircraft control activities. Here we see that experience, personality, cognitive ability and job knowledge, among the pilots we tested, were important factors associated with other performance measures. The factors involving multi-tasking are the most important factors in CogScreen. Notably absent from these bivariate correlations are the measures of general intelligence, but this is most likely due to the narrow range of variance for the pilots that we observed. Finally, pilots' ability to perform on knowledge-based tests is associated with their ability to perform in training.

We believe that it is important to collect data on populations that are far less homogeneous so we

can get a better indicator of what role cognitive ability plays in CRM. Theoretically, it can be argued that because our sample was so well trained, so knowledgeable and so bright, cognitive ability plays less of a role here than it would among a broader population of pilots. Because these pilots know their jobs well and have been performing them for years, so much of what they do is now routine and "hard-wired", requiring less stress on cognitive capacity. On the other hand, it can be argued that cognitive ability plays an important role here primarily because it is the one dimension on which these pilots were not originally screened. Only by reproducing this study in a variety of populations can we build a model of how all of these factors relate to performance in the cockpit.

BIBLIOGRAPHY

Ackerman, P.L., Schneider, W. & Wickens, C.D., "Deciding the Existence of a Time-Sharing Ability: A Combined Methodological and Theoretical Approach." *Human Factors, 26*, 1984, pp. 71-82.

Chidester, T.R., Helmreich, R.L., Gregorich, S.E. & Geis, C.E., "Pilot Personality and Crew Coordination: Implications for Training and Selection," *International Journal of Aviation Psychology, 1*(1), 1991, pp. 25-44.

Chidester, T.R., Kanki, B.G., Foushee, H.C., Dickinson, C.L. & Bowles, S.V., "Personality Factors in Flight Operations: I. Leader Characteristics and Crew Performance in Full-Mission Air Transport Simulation", (NASA Tech Memorandum No. 102259), Moffett Field, CA, NASA-Ames Research Center, 1990.

Cooper, G.E., White, M.D., and Lauber, J.K. (Eds.), "Resource Management on the Flight Deck" (NASA Conference Publication No 2120; NTIS No. N80-22283), Moffett Field, CA, NASA-Ames Research Center, 1979.

Costa, Paul T., Jr., and McCrae, Robert R., *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI),* Psychological Assessment Resources, Inc., 1992.

Damos, Diane L., "Using Meta Analysis to Compare the Predictive Validity of Single- and Multiple-Task Measures to Flight Performance", *Human Factors, 35,* 1993, pp. 615-628.

Foushee, H.C. & Helmreich, R.L., "Group Interaction and Flight Crew performance." In E.D. Wiener & D.C. Negel (Eds.), *Human Factors in Aviation,* San Diego, CA, Academic, 1988, pp. 189-227.

Helmreich, R. & Spence, J., "The Work and Family Orientation Questionnaire: An Objective Instrument to Assess Components of Achievement Motivation and Attitudes toward Family and Career", *JSAS Catalog of Selected Documents in Psychology, 8*, 1978, p. 35.

Imhoff, D.L., & Levine, J.M., "Perceptual-motor and Cognitive Performance Task Battery for Pilot Selection", (AFHRL Tech. Rep. No. 87-20, AD-A094 317), Brooks Air Force Base, TX, Manpower & Personnel Division, Air Force Human Resources Laboratory, 1981.

Kay, Gary G., *CogScreen: Professional Manual*, Psychological Assessment Resources, Inc., Odessa, Florida, 1995.

Kay, G.G., Hordinsky, J.R., & Pakul, B., "Neuropsychological Assessment of Aviators: A Comparison of Traditional and Computer-based Approaches", Annual Meeting of the Aerospace Medical Association, 1992.

Mosier-O'Neill, K.L., "A Contextual Analysis of Pilot Decision Making." In R.S. Jensen (Ed.), *Proceedings of the Fifth International Symposium on Aviation Psychology,* Columbus, OH, Ohio State University, 1989.

O'Hare, David & Wiggins, Mark, "Expertise in Aeronautical Decision Making: A Cognitive Skill Analysis." In R.S. Jensen & D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology*, Columbus, OH, Ohio State University, 1993.

Olea, M.M. & Ree, M.J., "Predicting Pilot and Navigator Criteria: Not Much More Than *g*", *Journal of Applied Psychology*, *79*, 1994, pp. 845-951.

Pred, R.S., Spence, J.T. & Helmreich, R.L., "The Development of New Scales for the Jenkins Activity Survey Measure of the Type A Construct", *Social and Behavioral Science Documents (MS No. 2769), 16*, 1986, pp. 51-52.

Ree, M.J., and Carretta, T.R., "The Central Role of *g* in Military Pilot Selection", Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Research Division, 1995.

Spence, J., Helmreich, R.,& Holahan, C., "Negative and Positive Components of Psychological Masculinity and Femininity and their Relationships to Self reports of Neurotic and Acting-act Behaviors", *Journal of Personality and Social Psychology, 37,* (1979), pp. 1673-1682. [This may be the original validating study on the Personality Attributes Questionnaire.]

Yakimovich, N.V., Strongin, G.L., Govorushenko, V.V., Schroeder, D., & Kay, G.G., "Flight Performance and CogScreen Test Battery Performance in Russian Pilots", Annual Meeting of the Aerospace Medical Association, 1995.

Table 1

Factors	**Cronbach's Alpha**
Procedures Compliance and Checklists	.92
Crew Resource Management	.96
Aircraft Control Activities	.90
Planning and Preparation	.83

Figure 1

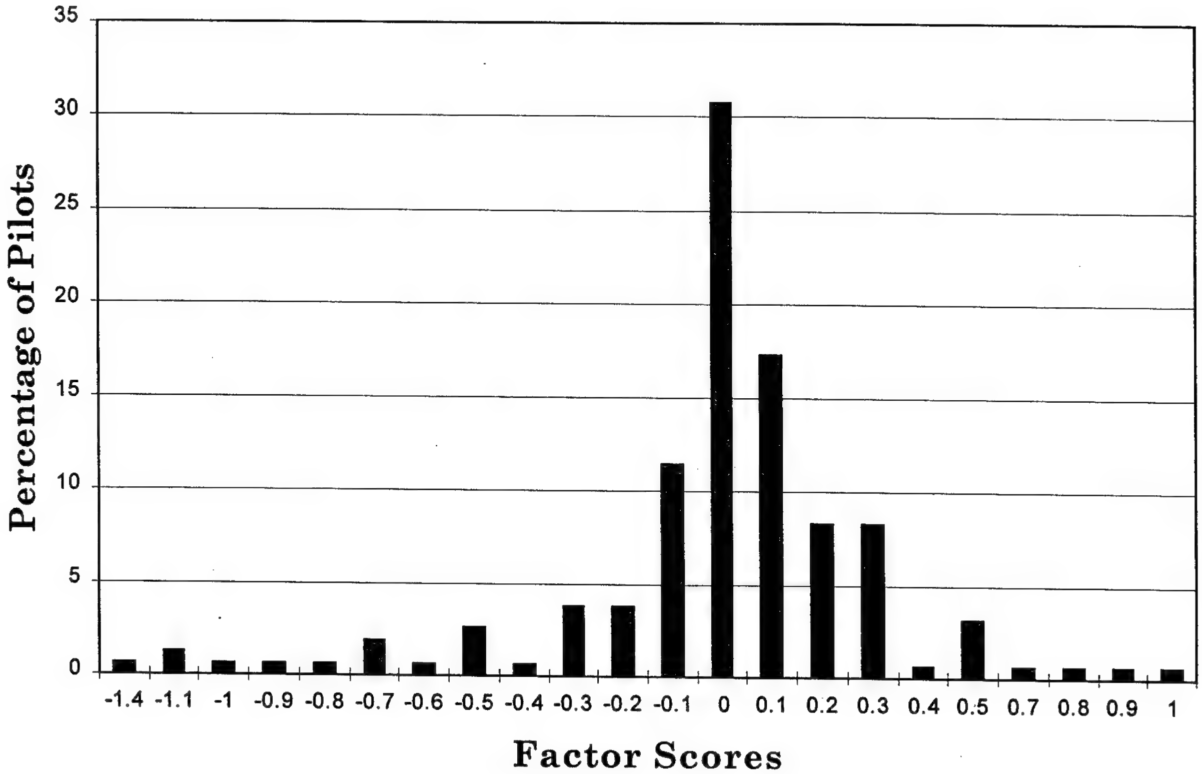

35
30
25
20
15
10
5
0
Percentage of Pilots
-1.4 -1.1 -1 -0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.7 0.8 0.9 1
Factor Scores

Table 2
Means and Standard Deviations of Measures with Significant Correlations with CRM

Measure	n	Mean	Standard Deviation
Backward Digit Span Accrcy	110	87.17	15.79
Math Accrcy	108	70.37	14.07
Math Speed	108	27.30	8.40
Math Thrpt	108	1.72	0.71
Visual Seq Comprsn Accrcy	110	98.45	2.69
Visual Seq Comprsn Speed	110	2.17	0.48
Visual Seq Comprsn Thrpt	110	28.52	5.96
Symbol Digit Coding Accrcy	110	99.36	1.28
Symbol Digit Coding Thrpt	110	34.67	5.74
Immediate Recall Accrcy	110	96.23	10.18
Delayed Recall Accrcy	110	92.31	18.80
Matching to sample Accrcy	110	95.73	4.68
Matching to sample Speed	110	1.36	0.24
Matching to sample Thrpt	110	43.56	7.67
Manikin test Accrcy	110	95.64	6.21
Manikin test Speed	110	1.47	0.32
Manikin test Thrpt	110	40.74	8.78
Div attn test Indctor alone speed	110	0.34	0.08
Div attn test Indctr alone prem resp	110	3.43	2.26
Div attn test Indctr dual speed	110	0.61	0.20
Div attn test Indctr dual prem resp	110	1.67	1.93
Div attn test Seq comprsn accrcy	110	95.11	7.27
Div attn test Seq comprsn speed	110	2.19	0.71
Div attn test Seq comprsn thrpt	110	28.37	8.04
Auditory seq comprsn Accrcy	109	97.06	5.49
Auditory seq comprsn Speed	109	0.61	0.14
Auditory Seq comprsn Thrpt	109	100.28	23.34
Pathfinder Number accrcy	109	99.53	1.54
Pathfinder Number speed	109	0.79	0.16
Pathfinder Number thrpt	109	78.77	15.29
Pathfinder Number coordination	109	1.15	0.54
Pathfinder Letter accrcy	110	99.35	2.02
Pathfinder Letter speed	110	0.77	0.39
Pathfinder Letter thrpt	110	82.75	15.55
Pathfinder Letter coordination	110	1.25	0.47
Pathfinder Combined accrcy	110	98.49	2.94
Pathfinder Combined speed	110	1.10	0.47
Pathfinder Combined thrpt	110	58.74	14.55
Pathfinder Combined coordination	110	1.17	0.42
Shftng attntn test Arrow dirctn accrcy	110	99.19	2.70
Shftng attntn test Arrow dirctn speed	110	0.53	0.09

Shftng attntn test Arrow dirctn thrpt	110	114.27	18.24
Shftng attntn test Arrow color accrcy	110	99.32	3.04
Shftng attntn test Arrow color speed	110	0.61	0.09
Shftng attntn test Arrow color thrpt	110	99.39	15.07
Shftng attntn test Instrctn accrcy	110	98.45	2.39
Shftng attntn test Instrctn speed	110	0.72	0.13
Shftng attntn test Instrctn thrpt	110	84.26	14.85
Shftng attntn test Discvry accrcy	110	65.12	14.12
Shftng attntn test Discvry speed	110	0.90	0.26
Shftng attntn test Discvry thrpt	110	47.01	15.39
Shftng attntn test Discvry rule shft cmp	110	6.77	2.71
Shftng attntn test Discvry failed set	110	1.82	1.60
Shftng attntn test Discvry persev err(s)	110	2.35	3.08
Shftng attntn test Discvry noncncpt resp	110	2.32	3.45
Dual task test Tracking alone error	108	16.67	14.38
Dual task test Tracking boundary hits	108	1.88	3.17
Dual task test Tracking dual error	109	53.49	25.07
Dual task test Tracking dual hits	109	2.48	2.91
Dual task test Prev num alone accrcy	109	92.81	6.59
Dual task test Prev num alone speed	109	0.50	0.26
Dual task test Prev num alone thrpt	109	138.97	63.31
Dual task test Prev num dual accrcy	108	87.44	11.53
Dual task test Prev num dual speed	108	0.57	0.26
Dual task test Prev num dual thrpt	108	118.75	74.27
Bennets number correct	109	58.32	5.00
DAT num number correct	110	20.88	2.77
DAT vrb number correct	110	28.10	1.98
DAT tot number correct	110	48.98	3.76
JKT tot number correct # Crrct-rescore	109	69.92	9.13
JKT Aerodynamics % Crrct-rescore	109	75.73	12.86
JKT Air Navigation % Crrct-rescore	109	70.02	12.69
JKT Engineering % Crrct-rescore	109	85.12	9.42
JKT Aviation Physiology % Crrct-rescore	109	67.58	27.39
JKT Meteorology % Crrct-rescore	109	64.80	27.33
JKT Airmans Info Manual % Crrct-rescore	109	28.90	33.54
MIN tot number correct	110	48.27	6.34
Ravens tot number correct	109	23.98	4.07
Watson tot number correct	109	32.74	4.17
Total interview score	110	30.15	3.86
Neuroticism	106	53.77	14.79
Extraversion	106	122.49	12.87
Openness	106	109.19	15.15
Agreeableness	106	126.75	15.39
Conscientiousness	106	138.33	13.77
Fighter, capt instr, trner instr, ME jet	110	0.45	0.50
Transport, capt instr, trner instr, ME jet	110	0.29	0.46

Table 3
Correlation of Predictor Measures with CRM

Measure	n	Correlation Coefficient	Significance Level
Backward Digit Span Accrcy	110	-.11	.25
Math Accrcy	108	.08	.44
Math Speed	108	-.09	.36
Math Thrpt	108	.07	.50
Visual Seq Comprsn Accrcy	110	.10	.31
Visual Seq Comprsn Speed	110	-.03	.72
Visual Seq Comprsn Thrpt	110	.07	.50
Symbol Digit Coding Accrcy	110	-.08	.42
Symbol Digit Coding Thrpt	110	.06	.54
Immediate Recall Accrcy	110	.11	.25
Delayed Recall Accrcy	110	-.12	.21
Matching to sample Accrcy	110	.20	.03
Matching to sample Speed	110	.03	.72
Matching to sample Thrpt	110	.02	.84
Manikin test Accrcy	110	-.05	.57
Manikin test Speed	110	-0.17	.08
Manikin test Thrpt	110	.12	.20
Div attn test Indctor alone speed	110	.007	.94
Div attn test Indctr alone prem resp	110	-.05	.58
Div attn test Indctr dual speed	110	.04	.70
Div attn test Indctr dual prem resp	110	.07	.48
Div attn test Seq comprsn accrcy	110	.08	.40
Div attn test Seq comprsn speed	110	-0.16	.10
Div attn test Seq comprsn thrpt	110	.18	.06
Auditory seq comprsn Accrcy	109	-.05	.62
Auditory seq comprsn Speed	109	.04	.67
Auditory Seq comprsn Thrpt	109	-.04	.70
Pathfinder Number accrcy	109	.03	.79
Pathfinder Number speed	109	-.11	.28
Pathfinder Number thrpt	109	.09	.38
Pathfinder Number coordination	109	-.04	.65
Pathfinder Letter accrcy	110	.02	.82
Pathfinder Letter speed	110	-.15	.13
Pathfinder Letter thrpt	110	.02	.80
Pathfinder Letter coordination	110	.08	.39
Pathfinder Combined accrcy	110	-.10	.28
Pathfinder Combined speed	110	-.10	.30
Pathfinder Combined thrpt	110	-.02	.84
Pathfinder Combined coordination	110	.08	.38
Shftng attntn test Arrow dirctn accrcy	110	-.06	.56
Shftng attntn test Arrow dirctn speed	110	-.09	.34
Shftng attntn test Arrow dirctn thrpt	110	.07	.50

Shftng attntn test Arrow color accrcy	110	.10	.29
Shftng attntn test Arrow color speed	110	-.17	.08
Shftng attntn test Arrow color thrpt	110	.16	.09
Shftng attntn test Instrctn accrcy	110	.11	.24
Shftng attntn test Instrctn speed	110	-.13	.19
Shftng attntn test Instrctn thrpt	110	.12	.20
Shftng attntn test Discvry accrcy	110	.16	.10
Shftng attntn test Discvry speed	110	-.002	.98
Shftng attntn test Discvry thrpt	110	.12	.22
Shftng attntn test Discvry rule shft cmp	110	.19	.05
Shftng attntn test Discvry failed set	110	-.18	.05
Shftng attntn test Discvry persev err(s)	110	.07	.44
Shftng attntn test Discvry noncncpt resp	110	-.11	.26
Dual task test Tracking alone error	108	.003	.97
Dual task test Tracking boundary hits	108	-.07	.48
Dual task test Tracking dual error	109	.20	.03
Dual task test Tracking dual hits	109	.16	.09
Dual task test Prev num alone accrcy	109	-.03	.78
Dual task test Prev num alone speed	109	.10	.32
Dual task test Prev num alone thrpt	109	-.11	.24
Dual task test Prev num dual accrcy	108	.04	.66
Dual task test Prev num dual speed	108	-.09	.34
Dual task test Prev num dual thrpt	108	.12	.86
Bennets number correct	109	-.01	.90
DAT num number correct	110	-.06	.52
DAT vrb number correct	110	.02	.87
DAT tot number correct	110	-.04	.70
JKT tot number correct # Crrct-rescore	109	.09	.33
JKT Aerodynamics % Crrct-rescore	109	.13	.19
JKT Air Navigation % Crrct-rescore	109	.06	.57
JKT Engineering % Crrct-rescore	109	.006	.94
JKT Aviation Physiology % Crrct-rescore	109	.03	.78
JKT Meteorology % Crrct-rescore	109	.01	.88
JKT Airmans Info Manual % Crrct-rescore	109	.04	.66
MIN tot number correct	110	-.08	.42
Ravans tot number correct	109	.14	.13
Watson tot number correct	109	-.04	.70
Total interview score	110	.05	.58
Neuroticism	106	-.13	.17
Extraversion	106	.08	.40
Openness	106	-.09	.35
Agreeableness	106	.25	.009
Fighter, capt instr, trner instr, ME jet	110	.21	.03
Transport, capt instr, trner instr, ME jet	110	-.31	.001

Table 4

Regression Analysis of CRM Model $R^2 = .353$

Variable	Parameter Estimate	Significance Level
Transprt, capt instr, trnr instr, ME jet	-0.3609	0.0001
(A) Agreeableness	0.0081	0.0020
JKT Aerodynamics % Crrct-rescore	0.0092	0.0051
Matching to sample Accrcy	0.0257	0.0025
Div attn test Seq comprsn speed	-0.1231	0.0278
Dual task test Tracking dual hits	0.0304	0.0239
DAT tot number correct	-0.0273	0.0182
Intercept	-2.4685	0.0115

Table 5
Significance of Bivariate Correlations at .1 Level

Tests of Ability and Aptitude	Training	Procedural Compliance	Aircraft Control	CRM
Numerical Ability				
Verbal Ability				
Mechanical Ability				
Spatial Ability				
Test of Reasoning				
Nonverbal Aptitude				

Personality	Training	Procedural Compliance	Aircraft Control	CRM
Neuroticism				
Openness				
Agreeableness		X		X
Conscientiousness	X			
Extroversion				

Knowledge-Based Tests	Training	Procedural Compliance	Aircraft Control	CRM
Knowledge of Aerodynamics	X			
Knowledge of Engineering	X			
Knowledge of Navigation				
Knowledge of Meteorology		X		
Knowledge of Aviation Physiology	X			

Cognitive Processing	Training	Procedural Compliance	Aircraft Control	CRM
Math Accuracy				
Visual Sequencing	X		X	
Symbol Digit Coding			X	
Matching to Sample	X	X		X
The Manikin Test		X	X	X
Divided Attention	X	X		X
Auditory Sequence Comparison				
Path Finder	X		X	
Shifting Attention	X	X	X	X
Dual Tasking	X	X		X

Prior Experience	Training	Procedural Compliance	Aircraft Control	CRM
Fighter experience		X	X	X
Large jet air transport		X	X	X
Turbo prop				

ACKNOWLEDGEMENTS

Our sincere thanks and appreciation to:

Sally H. Spetz, Ph.D.
April Sansom

APPENDIX A

CRM

Item		Factor Loading
16.4	Establishes and reinforces two-way communication	0.63
16.2	Provides feedback/accepts critique	0.61
21.7	Contributes proactively to the selection of a course of action	0.60
21.8	Accepts and executes decisions when finalized	0.59
16.5	Asserts perspective safety and/or efficiency	0.56
17.4	Informs captain of task progress and status	0.56
17.3	Resolves disparities in interpretation, priority and technique	0.55
21.4	Reviews assumptions and decisions before selecting a course of action	0.53
21.1	Looks for multiple cues to identify the problem	0.52
17.1	Backs up other crew members	0.51
18.2	Adapts to changes	0.51
19.1	Prioritizes individual tasks	0.51
21.3	Contributes proactively to the research of options	0.50
16.1	Listens actively	0.49
20.2	Ensures that distractions do not detract from overall crew situational awareness	0.49
18.1	Plans ahead	0.49
8.7	Maintains heading/navigation, altitude, airspeed tolerances	0.45
18.3	Executes plans as briefed	0.45
16.6	Asks for clarification when necessary	0.44
20.3	Maintains automation mode awareness	0.44
17.2	Discloses mistakes and/or limitations promptly	0.43
20.1	Maintains situational awareness throughout flight	0.43
21.5	Considers operational priorities and risk when selecting a course of action	0.42
19.2	Prepares for high workload during low workload	0.41
19.4	Uses the appropriate automation level for reducing workload	0.41
21.2	States symptoms, not conclusions, when initially identifying the problem	0.40
16.3	Uses standard terminology	0.38
14.2	Monitors autoflight systems for proper flight path control and performance	0.35
21.6	Considers time restraints when selecting a course of action	0.33

Training Group Performance for Biasing and Debiasing Decision Making which Avoids Groupthink.

Malcolm Cook and Leona Elder
Division of Psychology
School of Social and Health Sciences
University of Abertay Dundee
158 Marketgait
Dundee, DD1 1NJ, United Kingdom
Tel. 01382 308749 - Fx: 01382 223121
Answerphone : 01382 308709

George Ward
ESE Associates
15, Jesse Close
Yately, Hampshire, GU46 6AH
United Kingdom
e-mail : m.cook@river.tay.ac.uk

Summary

Research on biases in thinking and judgement are frequently related to the strategic use of limited information processing resources by human operators. Human operators have been shown to have a limited capacity short-term memory and to experience problems in retrieving information quickly from long-term memory. The limited information processing capability of the human operator is supposed to encourage the use of heuristics and biases which reduce memory requirements of processing (Huey and Wickens, 1993). Application of this model to decision making by operators in complex systems suggests that external cognitive support and effective information presentation are appropriate responses to increase the probability of correct decisions. In this paper it is argued that the reluctance to shift out of skill-based processing encourages the maintenance of biases in thinking. It is suggested that awareness of their own biases and of the periods in which they are likely to occur may render decision makers more effective. In addition, it suggests a new style of pilot's assistant technology which actively encourages the exchange of information between on-board systems and the operator. This participative dialogue management will help to ensure that inconsistencies between information and action are addressed before an ineffective mental model activated and applied to key decisions.

Introduction

Groupthink is a faulty decision making process that takes place when certain conditions define the local environment. Among those conditions, Forsyth (1990) has identified the following as most significant:

Pressure to Conform - groups seek consensual views.

Self-censorship - personal views which contradict or criticize are not expressed.

Mindguarding - information is pre-filtered according to a dominant viewpoint or to accord with perceived role responsibilities.

Apparent unanimity - public views support general agreement.

Illusions of invulnerability - the perceived threats or challenges are underestimated.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

Defective decision-making strategies - poor response evaluation or response selection processes.

Groupthink is fostered by cohesiveness between group members, isolation from other sources of information and leadership which manages the interactions tightly. There is often a degree of decisional stress associated with Groupthink and group members may feel a need to select a course of action quickly to reduce this stress. It can be argued that training regimes and poor systems design can promote these conditions and attitudes which are likely to foster Groupthink in multi-participant systems commonly found in military systems.

Biases in Thinking and Judgement

In a review, Evans (1989) has identified a large number of biases which can potentially limit the effectiveness of decision making in terms of judgemental or reasoning processes. Like other researchers (e.g. Anderson, 1990), Evans dealt with decision making from an individual point of view and did not discuss the interactions between group and individual cognitions. Stasson, Ono, Zimmerman and Davis (1988) found that groups can be subject to biases similar to those that occur in individuals. Their data indicate that both truth-supported performance improvements and bias induced decrements in decision making can occur in groups. Stasser (1988) used computer simulation to identify many of the possible mechanisms by which groups may fail to achieve an effective decision and attempted to model performance from experimental data. He suggested that groups frequently fail to discover that their collective knowledge favours one alternative and their individually held information supports another alternative. Modelling suggests that this failure can occur even if members do not bias their contributions to discussion in such a way as to support their preferences. Other evidence suggests that decision makers frequently pursue information that will fail to discriminate between competing hypotheses and will simply confirm their expectations (Baron, Beattie and Hershey, 1988; Gilhooly, 1983).

Discrepancies between group decision making and models of the decision making process are frequently explained by invoking human bias (Pete, Pattipati and Kleinman, 1993). There are two sources of bias, group and individual. It is clear that groups may introduce additional factors such as social influence into the decision making process For instance, in groups, self-serving biases may reinforce any tendency towards Groupthink and towards the development of the perception of unrealistic actions as necessary. Many complex systems have multiple participants who may potentially exchange knowledge that shapes that decision making process. Lim and Benbasat (1997) suggest that the same biases found in individuals, individuals are prone to appear in group decision making systems because they are attributable to the cognitive limitations experienced by the members of the group. Thus, it might be argued that individual biases remain intact in the face of communication.

However, many of the participants in the reported studies are not trained decision makers with formal instruction in decision making. Rogalski and Samurcay (1993) analysed communication in complex distributed decision making and found that effective functioning was linked to a well-structured flow of communication and role distribution. It is likely then, that many studies carried out with students may not provide adequate role definition which may interact strongly with the inadequate flow of communication in the groups tested. Although not intentional, the ineffective transmission of information may act upon group situational awareness in the same manner as mindguarding and self-censorship in promoting apparent agreement across the group.

There is evidence that biases can be observed as phenomena outside the laboratory in everyday behaviour (Gilovich, 1991). People show general tendencies to impute order to random phenomena, to refine the central point of stories omitting important qualifications and situational details and they frequently recall only positive instances, not non-confirming negative ones. Thus, there is a tendency to fail to store or recall cues which would contradict an incorrect reading of a developing situation. Biases in processing have been used to explain the difference between optimal decision making

performance and observed performance (Miao, Luh, Kleinman, and Castanon, 1991). While biases were originally proposed in early research based on mathematical models of decision making in textbooks (e.g. Bell, Raiffa and Tversky, 1988), the same biases can be related to alternative naturalistic decision making models proposed more recently (Christensen-Szalanski, 1993).

The Requirement for Effective Decision Making

One of the key incidents that has come under scrutiny in the military decision making literature is the USS Vincennes incident in which an American warship shot down an Iranian Airbus over the straits of Hormuz on 3rd July 1988. The crew of the USS Vincennes had suspected that the plane was an Iranian F-14 and believed incorrectly that it was descending in preparation for an attack (Brookes, 1996). This incident provided evidence of confirmation bias in that the letters F-14 scrawled on the chart in the combat-information-centre (CIC) and the reported flight path were in agreement with the hypothesis that the aircraft was a likely aggressor (Huey and Wickens, 1993). In addition the slow speed of the aircraft, below what might be expected for an attack, was rationalised as an attempt to confuse the operators and produce a reticent response. It was claimed that, in weeks before the incident, military aircraft had flown in close proximity to civil aircraft to avoid detection by radar. This had helped to fuel suspicion on this occasion. It seems that the crew of the warship had failed to keep the commanding officer completely or correctly informed with the result that an inappropriate decision was made.

Communication is possibly one of the first casualties of an incident in a multi-operator system because individual operators will seek to review their own assigned task domains to provide accurate and up-to-date information on the current state of the system and the operating environment. In an attempt to reduce unnecessary communication, operators seek to demonstrate competence by filtering out individual information but this may result in poor situational awareness. Interpretation of the situation and the perception of others views, may further bias this process in favour of one hypothesis by active searches for confirmatory information and denial of contradictory evidence.

Many real-time systems simply do not afford an opportunity for overall review of the situation and it is probable that operators will select the most likely interpretation without cross-validation against available information. The glass cockpit, for instance, is a reflection of the belief that more effective information display and increased automation will improve performance in complex systems. However, automation and advanced display systems in the form of a glass cockpit can paradoxically both make more information available and encourage less effective processing of the information by individual operators (Mosier and Skitka, 1996). In other words, while automation is intended to reduce operator workload and afford opportunity to review available information, it may in fact subtly change the patterns of communication. Automation induced changes in communication may, in turn, undermine performance by giving rise to conditions that would promote Groupthink. For example, Bowers et al. (1995) found that automation was not associated with better performance and that its presence negatively impacted the level of active involvement demonstrated by crew members. Some authors have argued that the trend towards more minimalistic intervention by human operators should not be viewed as threatening because despite there being a shift towards supervisory and monitoring roles for human operators they still interact with the systems (Wieringa.and van-Wijk, 1997). However, evidence from accident reports in aircraft, nuclear power systems and other complex systems suggests that some forms of limited interaction are implicated in precipitating catastrophic events from trivial events that would have been detected in less advanced systems. In the absence of empirical data, one might argue that the trend towards a supervisory and monitoring role is a poor match to human skills which are based on highly interactive manual involvement (Hollnagel, 1993).

Falling involvement is not restricted to single operators but also affects groups. Those groups using advanced automation typically spend more

time considering the information presented and less time discussing the information with their colleagues in multi-operator systems. Such disengagement from the task and roll-off in interactive critiquing is likely to encourage greater biases to develop. Billings (1997) has argued that automation must be comprehensible, must not remove operators from the command role and the primary role is in maintaining situational awareness. In some cases, accidents are attributed to a failure of systems to present appropriate information to operator, which lack of information then leads on to erroneous decision making (Martensson , 1995). Reduced interaction and unclear information presentation both contribute directly to poor decision making by eroding situational awareness. Effective multi-operator systems which require negotiated decision making clearly must make information available in a form that is easily understood and shared between participants. However, they t they may need to be designed specifically for the active promotion of information exchange to ensure an effective challenge to the development of bias. Since Groupthink develops in isolated groups with apparently consensual views when they are put under extreme pressure to make critical decisions, promotion of information exchange may prevent the development of such premature unanimity and ensure the maintenance of viable alternative hypotheses across the group.

It is possible, then, that specific design philosophies increase the likelihood of the development of poor situational awareness by encouraging disengagement, poor information presentation and reduced crew interactions. Compounding the problems inherent in the system design are other more variable influences such as fatigue. Many complex systems operate around the clock and over long periods of time. Yet, It is the case that frequency of operator interaction decreases with fatigue. It is also known that working memory capacity changes across the circadian period and can result in poor performance in relation to complex tasks (Proctor and Dutta, 1993). The decreasing capacity to manage information on the part of the individual operator may generate faulty perceptions of events such that unrepresentative information is seized on and retained. This information may go unchallenged because of the reduced intensity of crew co-ordination in fatigued states. In short, fatigue, sleep deprivation and disruption which are common in military operations may interact with aspects of the design to exacerbate problems in multi-crew systems which produce fewer and less effective crew interactions and poorly informed and less effective decisions.

Stress and Decision Making

The tendency to use biased processing, particularly under time pressure has been examined in the laboratory by Lehner, Seyed-Solorforough, O'Connor, Sak, and Mullin (1997). Lehner et al asked subjects to make judgements that were inconsistent with normal heuristic decision processing. The decision procedures were found to be vulnerable both to bias and to the effects of time stress. It was suggested that the decision makers in the two-man teams adapted inappropriately to time stress in that, as time stress increased, they began to use a decision processing strategy that was less effective than the strategy they had been trained to use. Thus, any training which is intended to produce more effective decision making must ensure that the operator is aware of the possible effects of time stress and the nature of the normal decision making process (Orasanu, 1993; Cannon-Bowers and Bell, 1997). Research on decision making as part of training suggests that decision making training is not simply about observing relationships but acting upon the decisions made (Berry, 1991).

In examining flawed decision making, Huey and Wickens (1993), have accepted that military accidents can be interpreted as support for the appearance of biases in decision making under the time pressure, high stress and high workloads which typically occur in military and civil cockpits. It is likely that such erroneous decision making will not be restricted to cockpits but iwill appear in communication, command, control and intelligence situations. The key elements of the problems are the diverse sources of information which must be integrated to produce effective decision making combined with the limited capacity of the human operator to form new plans from raw information. The present authors suggest that while operator's actions and plans at the micro and macro level are guided by the perception of the current

situation, the relationship between the macro and micro level can be de-coupled to allow apparently minor inconsistencies to be processed without changing the overall plan. It has been argued that inconsistency between the current working hypothesis and information may paradoxically strengthen first impressions. Errors and biases in decision making are arguably likely to arise when operators lose track of the number of inconsistencies they have processed at a micro level. However, it might be the case that one positive function of automation would the creation of a log of significant information exceptions that do not fit with the current view or the operation of query-based processes to cross-check the evidence in support of a current hypothesis. With the addition of time stamps or spatial patterns, these tools might be useful as training aids which can help operators develop a knowledge of the significance of cues in key decisions and improve their decision making under stress.

Experience and Information Usage

Huey and Wickens (1993) observed that many heuristics and biases were a result of human information processing capacity limits. Specifically, the limited capacity working memory, slow and problematic retrieval from long term memory, limited attentional capacity and the inability to place keep information in time were identified as major problems. Their suggestions are supported by Randel, Pugh and Reed's (1996) study of situation awareness as part of a decision making task in a naturalistic setting involving a complex cognitive task. Using twenty-eight electronic warfare technicians from U.S. Navy ships classified either as novices, intermediates, or experts, they showed that expertise includes proficiency in visually and verbally recalling radar emitters, the ability to make correct decisions based on better situation awareness, and the ability to understand the conditions for applying rules in a consistent manner. The final two skills are precisely the same forms of expertise observed in expert chess players.

Prior experience has both negative and positive aspects. It undoubtedly helps individuals to encode data more effectively and to retrieve that information. However, it can, conversely, undermine effective processing of novel information if the operator relies on seeking confirmatory evidence in support of a conclusion selected at an early stage in unfolding events. Research has certainly shown that experts spend more time encoding information correctly in preparation for decision making and response selection. Yet, the most significant failure in many critical events may derive from the operator's limited ability to bring to mind all the relevant information that is distributed in space and time despite the fact that prior experience provides extended practice of the rule-set and the opportunity of developing automaticity to free resources for difficult judgements. McKinney (1993) has provided evidence from an analysis of decision making after mechanical breakdowns to suggest that experience may indeed be a handicap on some occasions. He found that lead pilots with more experience made poorer decisions than relatively inexperienced wingmen. Huey and Wickens (1993) have also reported a study in which less experienced operators made more effective decisions in response to unusual events than did more experienced operators. They suggest that might be due to a greater reliance on available information by the less experienced while more experienced operators may either try to fit the pattern of events into prior experiences and discount contradictory information or may simply fail to realise its significance because they have not processed it.

It can be argued that that experienced operators may not use exhaustive searches of long-term memory to identify appropriate behaviours but opt instead for cued-recall or retrieval of strategies found to be optimal in the past. In support of this view, there is evidence that experts use less time and fewer steps in searches, employ different search strategies, and search a different amount of information than novices (Salterio, 1996). The different use of information by experts and novices has encouraged some designers to focus on information presentation as the route to more effective decision making in interfaces for supervisory control systems. For example, Coury and Semmel (1996) suggest that an intelligent interface would be more able to direct attention towards pertinent information and so enable more effective operator intervention. Many systems are triggered by significant events

which require prompt action and it might be argued that during critical events operators are less likely to incorporate new information unless it is both obviously salient and appropriate. The treatment of information may depend on the global picture perceived by the operator and new information which is at odds with the current working hypothesis requires additional processing to integrate it or revise the current mental model. It is difficult to refute the argument for the impact of limited cognitive resources on processing of information and the added complication of time stress on decision making, which makes significant information . However, it is practically very difficult to automatically isolate the significance of individual items of information, using intelligent agents embedded in the systems interface, and it may be that role which humans are more capable of.

In simple terms, operators under time stress are more likely to emit well learned responses in an effort to free capacity to allow an effective evaluation of future actions and plans. Simple mistakes in the aeronautical world such as attempting to land without gear or flying too low during normal flight have been suitable systems for this protective approach. However, even simple systems do not prevent *simple* mistakes and controlled flight into terrain remains a major problem for civil and military pilots with significant numbers of crashes each year. Often the difficulty with dynamic systems is with respect to looking ahead and predicting what knowledge will be required to respond to future events. It is all too obvious from cockpit voice recorders that pilots know what will happen next at a point when it is too late to apply any corrective action.

Place and Time

It may be that problems in place-keeping significant information and its temporal relationship to on-going events is equally important in the evaluation of unfolding events. If additional burdens are added to operators in those circumstances they are likely to reject, lose or fail to retrieve pertinent information that would guide their response to external demands. On the other hand, operators may simply choose not to process the information because of the high demand on limited resources. This type of failure can occur with respect to aeronautical decision making when weather, time and mechanical problems conspire to produce a lethal framework of circumstances which must be responded to sharply. The additional cognitive burdens and the stress of such circumstances can paradoxically reduce working memory capacity and attentional focus to the point where decision makers experience cognitive lockup. Kerstholt, Passenier, Houttuin and Schuffel (1996) found that operators given vigilance and monitoring roles along with remediation roles requiring intervention were susceptible to task shedding and inappropriate focus on single events. In teams, the first task to be sacrificed to time stress is communication and, as already argued, this is likely to reduce the effectiveness of decision making. The impact on military decision is likely to exaggerated by the loss of effective since teams are frequently distributed and require effective coordination (Fischhoff and Johnson, 1997). This suggests that teams and the systems they operate should have guaranteed distributed access to the information necessary for effective decisions. Participative interaction in such distributed information systems is likely to be more effective since passive observations of plans may not encourage effective plan reviews.

In Kersholt et al.'s (1996) work, subjects were observed first to ignore the monitoring function of their role when diagnosing a disturbance and then, as the probability of disturbances increased, to experience "cognitive lock-up" during which operators concentrated on single disturbances while ignoring the rest of the system and other tasks. It is equally likely that individuals will concentrate on activities related to their own situational awareness to the detriment of those communication activities which could maintain an effective team mental model (Urban, Weaver, Bowers and Rhodenizer,1996). There is, after all, a general tendency for operators facing complex decisions to make a less thorough evaluation of the problem and to process information selectively(Timmermans and Vlek, 1992). This, in turn, may result in a poorer decision if it remains unchallenged.

Cognitive Gradients Against Knowledge-Based Processing

The irrational retention of biased and ineffective processing techniques may reflect self-knowledge about the available cognitive resources for storing and processing information. If the standard skill-based, rule-based and knowledge-based processing suggested by Rasmussen (1986) is considered, the operators attitude to changing their mode of processing can be inferred. Knowledge-based processing is usually associated with learning a new task. This is likely to be error prone in that it is cognitively resource intensive and operators may not have extensive domain knowledge. Rule-based processing is an intermediate level of operation in which the operator has acquired some knowledge of the relationships between external cues in the task domain and effective responses to those cues. The intermediate level of operation should require less effortful processing and allow the operator to spend more time monitoring performance. The most highly developed skills generally become highly automatic by the time they reach this level of operation. The operator will develop a significant capability to carry out simultaneous tasks which have reached a degree of automaticity or are over-learned during skill-based operation. Logically, experienced pilots and operators in command and control systems who are likely to operate in a skill-based mode will require to make a downward shift into rule or knowledge-based processing during unusual events or situations. This shift may require the shedding of tasks to free resources for reallocation to the most difficult task. In this more demanding mode, operators may, consequently, miss valuable information that would normally inform their overall judgement processes or they may no longer have sufficient capacity necessary to detect further significant events.

Kaempf, Klein, Thordsen and Wolf (1996) examined the decision making processes of experienced naval officers in a complex, time-pressured command-and-control setting within the Combat Information Center of AEGIS naval cruisers. Their analysis of critical events suggests that the decision processes were consistent with a Recognition Primed Detection model and behaviour was predicated on recognition of the underlying trend in events. Operators were primarily concerned with situation awareness, and diagnostic activities flowed from the knowledge generated by situational awareness building. They reported that feature-matching and story formulation were used extensively and enabled the selection of appropriate actions. It is possible to construe this method of operations as a skilled set of cognitive-perceptual processes which automatically generate behavioural requirements for successful completion of current and future tasks. The operator is not generally aware of the long-term value of actions but is primarily monitoring the process to ensure that cognitive resources are not overwhelmed by demands nor key areas of performance undermined. The operator's focus on the process may help to explain why tasks requiring novel integration of diverse information sources are likely to produce fail. Few systems enable operators in command and control to track vertically and horizontally or to identify related information sources in separate sensing and display systems. Displays, themselves, frequently fail to fuse or inter-relate relevant information.

It is equally possible that operators become fixed on the goals or the outcomes of their task with important consequences for their interpretation of any information indicative of future hazards. It has been argued by Evans, Over and Manktelow (1993) that human decision makers frequently use a specific type of reasoning which focuses on achieving goals within certain cognitive constraints and which induces predictable errors. In addition, Evans and Over (1996) suggest that human decision makers are successful in making decisions related to achieving goals but relatively poor are generating inferences according to logic or probabilistic assumptions. O'Hare and Smitheram (1995) note that pilots fixated on the gains associated with actions tend to make risky and ineffective decisions. This could be re-interpreted as a focus on goals and a denial of probability. Investigations of crashes have indicated that pilots are equally likely to take risky decisions and to ignore the prevailing evidence against their chosen plan of action.

It is predicted that operators will be reluctant to shift from skill-based processing because there is

a cognitive resource demand associated with such a change. In multi-tasking during operation of complex systems, operators frequently rely upon automaticity across the full range of tasks to be able to accomplish the functional requirement of the total system. Shedding of tasks required by changes in processing mode in order to complete specific tasks which are more resource intensive, will generate reductions in overall performance and may jeopardise mission roles through a failure to produce an integrated response. The cognitive resource incline, perceived in relation to changes between skill-, rule- and knowledge-based processing of specific tasks, may, arguably, discourage operators from changing their current processing mode. Operators may tend to revert to heuristic and biased processing under time stress simply because automaticity allows concurrent tasks to be executed. By maintaining effective dialogue and situational awareness prior to significant events, therefore, this is more likely to be an effective response in respect of key decisions.

It is clear that current trends towards down sizing, reduced training and the introduction of sophisticated avionics may actually increase the likelihood of biases and a consequent increase in flawed performance. Automation, using intelligent agents or simple machine-intelligence, has often increased the individual operator's cognitive load and decreased the accessibility of information (Woods, Johannessen, Cook and Sarter, 1994) which discourages crew interaction. It can be argued that the increasing isolation of individual operators and the demands on their limited cognitive resources has decreased communication, a valuable activity in critiquing operators' situational awareness. It is well known that communication patterns change under increased demand but it need not directly have detrimental effects unless a specific situation arises where the shared information would have been critical. It has been found that frequency of communication and type of communication have been independently correlated with effective team performance. However, very few systematic analyses have directly associated communication activities with performance outcomes by process tracing decision making or information availability related to communication activities.

With reduced flight hours and reduced training times, reductions in the rates of accrual of expertise in novice pilots and fading of critical skills in qualified pilots may occur. It could be argued that it is not simply the total flight time, spent in training, which determines the effectiveness of pre-mission training and mission critical decision making but it is the proportion of the flight time in different tasks. There no evidence suggesting the homogeneity cognitive skills required for learning of specific tasks and even less on the effectiveness of carrying out part and whole-task training in relation to final mission performance. It is important to note that not all flying hours are combat related and pilots may experience very complicated scenarios infrequently. Even if the decision making capabilities are not eroded the appropriate behavioural profiles of training they may not be maintained sufficiently to foster effective co-ordination and teamwork. These reductions in training may interact with new shared display systems to foster unwarranted and unspoken assumptions in pilot's behaviour which may be heightened by operational demands. Where high demand exists from on-board systems and related tasks, communication is often the easiest and most natural of activities to reduce or distort. It was found that in civil aircraft, after crew reductions and the introduction of more automation, and as a consequence of the apparent reduced need to communicate, communication levels fell and gave rise to significant falls in situational awareness.

In effect team performance may be compromised by the development of biases because of incomplete, inaccurate or flawed information. This paper suggests methods for encouraging effective team-based decision making as a protection against development of biased judgement or reasoning. De-biasing of problem solving is seen as essential to the development of effective team-performance which,in turn, is supported by effective communication. Roth (1997b) has suggested that operators in complex team decision environments should receive training in cognitive skills and better problem representation in order to improve their decision making. Team and individual performance depends on an awareness of the likely problems and appropriate process control strategies to

minimise the risk of their appearance or impact. Catastrophic accident development normally relies on a sequence of events. Effective review and monitoring procedures may help to prevent such pathogenic developments in team and system operations.

Conclusions

Operators in multi-operator distributed systems should clearly experience a wide range of training scenarios and make extensive use of after action reviews to help in the knowledge building process (Cannon-Bowers and Bell, 1997). The development of appropriate domain knowledge should involve a thorough review of the cues used in problem analysis and response formulation to allow operators to self-critique the perceived situation. It is proposed that this critical review process will help overcome biases that may even appear in mental simulation used in problem solving (Klein and Crandall, 1995). Evans and Over (1997) argue that human beings can formulate decisions and explicitly reason in a manner that is broadly rational but accept that human decision makers do not tend to follow the predictions derived from theoretical descriptions of decision making) Even if it is accepted that human decision makers can follow instructions and deduce necessary conclusions in accordance with logical principles one must still determine that instruction intended to improve decision making under time-stress will have a positive effect on performance. Some would argue against the application of the traditional decision making and judgement literature to that which occurs in the control of real-time systems, in favour of recognition primed detection models like that proposed by Klein (1993).

Review of the present literature suggests that a failure to correct individual bias will result in the appearance of errors in groups decision making and that the tendency to error may be exacerbated by a number of factors which include ineffective communication or interaction, poor workload management, inappropriate use of automation and neglect of factors biasing decision making. In addition, it is argued that operators should be made aware of the likely effects of time stress on communication between group members and that they should be trained to optimise the exchange of information between team members. It is likely that the most effective preventative measures derive from broad-based training in which operators form adequate mental models of the team environment, the physical systems and the functional roles of the team members. Operators must be warned about the possible dangers of complacency and the need for self-review of processes and information.

Roth, 1997a would argue that time critical systems require planned responses to emergencies in order to proceduralise the response to critical events, and effective situational awareness to inform decision making. However, this strategy in itself will not guarantee protection against bias development. Critical reviews of individual and team roles should serve to increase an awareness of information needs and optimise information exchange. In time the degree of automaticity in the information exchange process should increase and operators should develop self-monitoring and review competence. The greatest danger is the development of a satisficing, when operators do the minimum necessary to sustain the process and release cognitive resources. The satisficing approach to tasks is more likely to induce biases or undermine situational awareness when non-participative automation is introduced as a means of reducing the cognitive burden of the operator and inadvertently results in disengagement. Implicit in this view is the recognition that the decision making process has an outcome that is predicated on an effective decision making process (Lipshitz, 1997) and that in turn is dependent on adequate information processing of a critical mass of relevant information in the period immediately prior to response execution.

It is clear that poor planning and preparation may result in additional cognitive burdens which cannot be met during the command and control operation of a real-time task. Human decision makers make a large variety of mistakes in both planning and decision making based on a small number of basic problems (Doerner and Schaub, 1994) but these may be resolved through training. One thing that has beecome clear is the relative poverty of the current technological

approaches using automation and advanced display systems. The human operator has had their current role shaped by the introduction of piecemeal automation which have left only the most difficult tasks for operators, such as decision making, while denying them a participative role which would make available the appropriate information, and as a result they are left to take the blame when things go wrong.

References

Anderson, N. H. (1990). A cognitive theory of judgement and decision. In N.H Anderson et al. Contributions to information integration theory, Vol. 1: Cognition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Baron, J., Beattie, J. and Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes*, 42, 1, pp 88-110.

Bell D. E., Raiffa, H. and Tversky A. (1988). *Decision making: Descriptive, normative, and prescriptive interactions.* Cambridge: Cambridge University Press.

Berry D.C. (1991) The role of action in implicit learning. *Quarterly Journal of Experimental Psychology*, 43A, 881-906.
Billings C.E. (1997) *Aviation Automation : The search for a human-centred approach.* Hove, U.K. : Lawrence Erlbaum Associates.

Bowers, C., Deaton, J., Oser, R., Prince, C. et-al. (1995). Impact of automation on aircrew communication and decision-making performance. *International Journal of Aviation Psychology*, 5, 2, pp 145-167

Brookes, A. (1996). *Disaster in the Air.* Shepperton: Ian Allan.

Cannon-Bowers, J.A. and Bell, H.H. (1997). Training decision makers for complex environments: Implications of the naturalistic decision making perspective. In C.E. Zsambok, G. Klein et-al. (Eds.) *Naturalistic decision making. Expertise: Research and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Christensen-Szalanski, J.J. (1993). A comment on applying experimental findings of cognitive biases to naturalistic environments. In G. A. Klein, J. Orasanu et al. (Eds.) *Decision making in action: Models and methods.* Norwood, NJ: Ablex Publishing Corp.

Coury, B.G. and Semmel R.D. (1996). Supervisory control and the design of intelligent user interfaces. In R. Parasuraman & M. Mouloua, (Eds.) *Automation and human performance: Theory and applications. Human factors in transportation.* Mahwah, NJ: Lawrence Erlbaum Associates.

Doerner D. and Schaub H. (1994). Errors in planning and decision-making and the nature of human information processing. *Applied Psychology: An International Review*, 43, 4, pp 433-453.

Evans, J., St.-B, T. (1989). *Bias in Human Reasoning: Causes and consequences.* Hove: Lawrence Erlbaum Associates.

Evans, J., St.-B, T. and Over, D.E. (1996). *Rationality and Reasoning.* Hove: Erlbaum (Uk)/Taylor and Francis Ltd.

Evans J.-St.-B.-T. and Over D.E. (1997) Rationality in reasoning: The problem of deductive competence.
SO: Cahiers-de-Psychologie-Cognitive/Current-Psychology-of-Cognition, 16,1-2, 3-38.

Evans, J., St.-B, T., Over, D.E. and Manktelow, K.I. (1993). Reasoning, decision making and rationality. *Cognition*, 49,1-2, pp165-187.

Forsyth, D.R. (1990). *Group Dynamics*, 2nd Ed. Califormia : Brooks/Cole Publishing.

Fischhoff, B. and Johnson, S. (1997). The possibility of distributed decision making. In Z. Shapira (Ed.). *Organizational Decision Making.* Cambridge Series on judgement and decision making. New York, NY: Cambridge University Press.

Gilhooly K. (1983) *Thinking.* London : Academic Press.

Gilovich, T. (1991). How We Know What Isn't So: The fallibility of human reason in everyday life. New York, NY: Free Press.

Huey B. and Wickens C.D. (1993) *Workload in Transition*. New York, U.S.A. : National Academy Press.

Hollnagel E. (1993) *Human Reliability Analysis : Context and Control.* London : Academic Press.
Kaempf, G.L., Klein, G., Thordsen, M.L. and Wolf,S. (1996). Decision making in complex naval command and control environments. *Human Factors*, 38, 2, pp 220-231.

Kerstholt J.H., Passenier, P.O., Houttuin, K. and Schuffel, H. (1996). The effect of a priori probability and complexity on decision making in a supervisory control task. *Human-Factors*, 38, 1, pp 65-78.

Klein G. (1993) Naturalistic Decision Making : Implications for Designs. Wright Patterson Air Force Base, Ohio, U.S.A. : Crew Systems Ergonomics Information Analysis Center.

Klein, G. and Crandall, B.W. (1995). The role of mental simulation in problem solving and decision making. In P. Hancock, J. M. Flach et al. (1995). Local Applications of the Ecological Approach to Human-Machine Systems, Vol. 2: Resources for ecological psychology. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lehner P., Seyed-Solorforough, M. M., O'Connor, M.F., Sak, S. and Mullin, T. (1997). Cognitive biases and time stress in team decision making. *IEEE-Transactions on Systems, Man and Cybernetics Part A: Systems and Humans*, 27, 5, pp 698-703.

Lim,L.H. and Benbasat, I. (1997). The debiasing role of group support systems: An experimental investigation of the representativeness bias. *International Journal of Human Computer Studies*, 47, 3,pp 453-471.

Lipshitz. R. (1997). Naturalistic decision making perspectives on decision errors. In C. E. Zsambok and G. Klein (Eds). *Naturalistic Decision Making. Expertise: Research and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Martensson, L. (1995). The aircraft crash at Gottroera: Experiences of the cockpit crew. *International Journal of Aviation Psychology*, 5, 3, pp 305-326

McKinney, E. H. (1993). Flight leads and crisis decision-making. *Aviation, Space, and Environmental Medicine*, 64, 5, pp 359-362.

Miao, X., Luh, P. B., Kleinman, D.L.and Castanon, D.A. (1991). Distributed stochastic resource allocation in teams. *IEEE Transactions on Systems:Man and Cybernetics*, 21, 1, pp 61-70.

Mosier, K.L. and Skitka, L.J. (1996). Human Decision Makers and Automated Decision Aids: Made for Each Other ? In R. Parasurman and M. Mouloua (Eds.) *Automation and Human Performamce : Theory and Applications*. New Jersey : Lawrence Erlbaum Associates.

O'Hare, D. and Smitheram,T. (1995). "Pressing on" into deteriorating conditions: An application of behavioral decision theory to pilot decision making. *International Journal of Aviation Psychology*, 5,4, pp 351-370.

Orasanu J. (1993) Decision-making in the cockpit. In E.L. Weiner, R.L. Helmreich and B.G. Kanki (Eds.) *Cockpit Resource Management.* London : Academic Press.

Pete, A., Pattipati, K.R. and Kleinman, D. L. (1993). Distributed detection in teams with partial information: A normative-descriptive model. IEEE Transactions on Systems: Man and Cybernetics, 23(6), pp 1626-1648.

Proctor R.W. and Dutta A. (1995) *Skill Acquisition and Human Performance*. London : Sage Publications.

Randel, J.M., Pugh, H.L. and Reed, S.K. (1996). Differences in expert and novice situation awareness in naturalistic decision making. *International Journal of Human Computer Studies*, 45, 5, pp 579-597.

Rasmussen J. (1986) *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. Amsterdam : North-Holland.

Rogalski, J. and Samurcay, R. (1993). Analysing communication in complex distributed decision-making. *Ergonomics*, 36, 11, pp 1329-1343.

Roth , E.M. (1997a). Analysis of decision making in nuclear power plant emergencies: An investigation of aided decision making. In Zsambok C. E. and Klein G. (Eds). Naturalistic decision making. Expertise: Research and applications. (pp. 175-182). Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.

Roth E.M. (1997b). Analyzing decision making in process control: Multidisciplinary approaches to understanding and aiding human performance in complex tasks. In C.E. Zsambok, G. Klein et al. (Eds.) *Naturalistic Decision Making. Expertise: Research and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Salterio, S. (1996). Decision support and information search in a complex environment: Evidence from archival data in auditing. *Human Factors,* 38, 3, pp 495-505 .

Stasser, G. (1988). Computer simulation as a research tool: The DISCUSS model of group decision making. *Journal of Experimental Social Psychology*, 24(5), pp 393-422.

Stasson, M.F., Ono, K., Zimmerman, S. K. and Davis, J.H. (1988). Group consensus processes on cognitive bias tasks: A social decision scheme approach. *Japanese Psychological Research*, 30, 2,
pp 68-77.

Timmermans, D. and Vlek, C. (1992). Multi-attribute decision support and complexity: An evaluation and process analysis of aided versus unaided decision making. *Acta Psychologica*, 80(1-3), pp 49-65.

Urban, J.M., Weaver, J.L., Bowers, C.A. and Rhodenizer, L. (1996). Effects of workload and structure on team processes and performance: Implications for complex team decision making. *Human Factors*, 38, 2, pp 300-310.

Wieringa, P. A.and van-Wijk, R.A. (1997). Operator support and supervisory control. In T.B. Sheridan, T. Van Lunteren et al. (1997). *Perspectives on the Human Controller: Essays in honor of Henk G. Stassen*. Mahwah, NJ: Lawrence Erlbaum Associates.

Woods D.D., Johannesen L.J., Cook R.I. and Sarter N.B. (1994) Behind Human Error: Cognitive Systems, Computers, and Hindsight. Wright Patterson Air Force Base, Ohio, U.S.A. : Crew Systems Ergonomics Information Analysis Center.

ENHANCING MULTI-CREW INFORMATION WARFARE PERFORMANCE: AN EVENT-BASED APPROACH FOR TRAINING

Randall L. Oser, Daniel J. Dwyer, Janis A. Cannon-Bowers, & Eduardo Salas
Naval Air Warfare Center Training Systems Division Code 4961
12350 Research Parkway
Orlando, FL, 32826-3275 USA

SUMMARY

Successful performance in most complex military environments includes the ability to conduct information-based warfare (IW). In general, IW refers to the effective management of information (i.e., acquire, analyze, disseminate, protect) to achieve tactical objectives. While technology is an important component of IW, recent military and peace-keeping operations have highlighted the importance of the human component to IW.

Increasingly, IW is performed by teams of teams. Often these multi-crew teams are geographically separated and composed of personnel from different functional areas, services, and/or countries. The teams must be able to effectively coordinate despite numerous differences (e.g., terminology, procedures, systems, language, cultural). Clearly these factors pose a considerable challenge for effective performance. Unfortunately there are few efforts to identify strategies or methods that can be used to support multi-crew IW training.

One framework that has recently demonstrated considerable potential for establishing training in multi-team environments is the Event-Based Approach for Training (EBAT). EBAT is a framework that: (a) structures training opportunities using appropriate methods, strategies, and tools, (b) tightly links critical tasks, learning objectives, exercise design, performance measurement, and feedback, and (c) has resulted in improved performance in team training environments.

This paper will describe the application of EBAT within the multi-crew IW context by: (a) providing an overview of IW, (b) presenting a conceptual model of a learning environment, (c) forwarding a detailed description of EBAT, (d) presenting an EBAT example for multi-crew IW training, and (e) discussing considerations for implementing multi-crew IW team training.

INTRODUCTION

Information-based warfare (IW) has always been a critical component of military operations (Ref 1). However, IW is becoming increasingly more complex due to technological advancements (e.g., sensors, networks, computers) (Ref 2). These advancements can provide increased IW capabilities for: (a) gaining access to enemy information, (b) denying access to information by the enemy, and (c) manipulating information used by the enemy.

While technology is clearly an important component of IW, the manner in which the technology is used by humans is also critical. Many IW tasks require planning, decision making, situation assessment, and resource management (Ref 3). These important tasks can be supported by technology but cannot be adequately performed by technology alone.

Most IW operations are performed by teams of teams that may be comprised of personnel from multiple functional areas, services, or nations (Ref 4). As a result, the teams often possess different procedures, terminology, tactics, systems, organizational structures, information requirements, and geographical locations. These factors complicate the ability for multi-crew IW teams to effectively perform.

The success of multi-crew IW teams will be dependent upon preparedness. An important component of preparedness involves the manner in which the teams are trained. Technological advancements in the areas of the models, simulations, and networks hold considerable promise for creating effective training environments (Ref 5).

While these advancements have resulted in new technologies with potential for scenario-based training, technology alone does not ensure that effective learning will result (Ref 6). Training technology must support the establishment of effective learning environments. Unfortunately, few frameworks exist to guide how to best use these systems to support the development of effective learning environments (Ref 7).

This paper will describe a systematic framework for conducting training within the multi-crew IW context. The following sections will: (a) provide an overview of IW, (b) present a conceptual model of a learning environment, (c) forward a detailed description of the framework for training, (d) present an example of the framework within a multi-crew IW team training context, and (e) discuss a set of considerations for conducting multi-crew IW team training.

Paper presented at the RTO HFM Symposium on "Collaborative Crew Performance in Complex Operational Systems", held in Edinburgh, United Kingdom, 20-22 April 1998, and published in RTO MP-4.

INFORMATION WARFARE OVERVIEW

Prior to discussing a framework for conducting multi-crew team training it is important to forward a definition of IW. Neilson and Giasson (Ref 8) define IW as "an approach to conflict focusing on the management and use of information in all its forms and at all levels to achieve a decisive advantage in pursuit of national security goals. Information-based Warfare is both offensive and defensive in nature – ranging from measures that prohibit adversaries from exploiting information to corresponding measures to assure the integrity, availability and interoperability of friendly information assets..." pp. 545. This definition suggests that IW is a complex construct.

Libicki (Ref 9) identifies seven different forms of the IW domain. These are: command and control warfare (C2W), intelligence-based warfare (IBW), electronic warfare (EW), psychological warfare (PSYW), hacker warfare, economic information warfare, and cyber warfare. While C2W, IBW, EW, and PSYW are traditional areas of IW, the other forms are becoming increasingly important (Ref 9). Although the general objective of each IW form is the same (e.g., information access, denial, protection), the specific methods used to accomplish IW in each of the areas differ. For example, C2W may involve attacks to deny information to an enemy's command structure, whereas PSYW might involve distributing information to de-moralize the civilian population. It should be noted that although each of the seven forms are listed separately, IW generally involves more than one of the forms being performed in a simultaneously manner.

A CONCEPTUAL MODEL OF THE LEARNING ENVIRONMENT

Effective performance of complex tasks does not happen by chance, it must be learned. As a result, it is critical to establish an environment where such learning can occur. Effective learning environments are systems that facilitate the ability of the training audience to develop and maintain the competencies (i.e., knowledges, skills, abilities, and attitudes) necessary to perform required tasks. Learning environments must employ systematic, deliberate approaches to achieve critical task requirements of the training audience.

Four important characteristics that need to be considered when establishing a learning environment include: who is being trained, what is being trained, under what conditions is it trained, and how should it be trained. One way to conceptualize the relationship between these factors is depicted in Figure 1. The next section will briefly describe each component of the learning environment model in more detail.

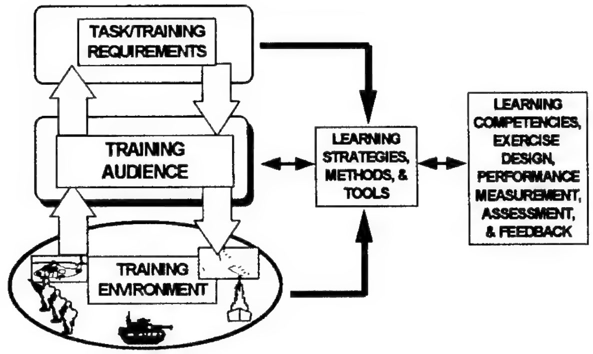


Figure 1. A Conceptual Model of a Learning Environment

Training Audience

Clearly, the training audience is at the center of the learning environment model. Because multi-crew IW is performed by teams, it is necessary to understand key characteristics of teams that will impact training. First, the extent to which members of a team possess an appropriate degree of shared understanding can significantly impact team performance in positive and negative ways (Ref 10). While multi-crew IW teams can bring a considerable level of expertise and resources to task performance, it is possible that the diversity can complicate performance. Because members from one team may not possess a detailed understanding about the systems, procedures, terminology, and tactics used by other teams, the establishment of the shared understanding may be difficult. The challenge to effective coordination is further complicated in the case of multi-national IW teams.

Second, the nature of a team's organizational structure can impact the requirement for and ability to exchange information (Ref 11). Because of the nature of IW, the multi-crew teams often possess complex organizational structures. The membership of the teams often includes a range of personnel from senior staff members responsible for the overall operation to junior and enlisted personnel responsible for sub-elements of the operation. The information required by senior staff members is different than that required by junior staff members and enlisted personnel. While IW teams are often organized in a hierarchical manner, the teams must be capable of adapting their structures in response to changing situations.

Third, the physical location of the team members can also impact the ability of the team to perform (Ref 12). Many multi-crew IW teams perform in distributed environments. The separation limits the team members' ability to coordinate using traditional cues. Recent advances in networking technologies have greatly improved the ability of geographically distributed teams to interact despite being physically separated. In most cases, coordination among

members of distributed teams must occur via technologically mediated means. A challenge for distributed teams is to ensure that interactions support effective coordination despite being geographically separated.

Task Requirements

Although the characteristics of the training audience are important, the requirements of the tasks to be performed by the training audience must also be understood (Ref 6). IW involves complex tasks associated with decision making, resource management, and situation assessment. In order to perform these tasks, multi-crew IW teams must be capable of accessing a considerable amount of information from a variety of sources. The teams must effectively perform despite information that may be incomplete, ambiguous, contradictory, or inaccurate. A challenge for these teams is to coordinate and synthesize the information in such a manner that IW can be effectively performed.

Multi-crew IW teams perform tasks that require immediate and future actions. Immediate actions are required to respond to time critical events (e.g., seconds or minutes). These tasks are often performed using rapid, or often automatic, responses. These responses will generally result in instantaneous feedback about performance. Tasks relating to future actions involve the development of longer term strategies and plans (e.g., hours, days, weeks, months). Strategy development and planning can rarely be accomplished using rapid and automatic responses. Feedback from these tasks is often delayed until the strategy or plan is actually implemented. Regardless of whether the feedback is immediate or delayed, the teams must use the feedback to effectively modify plans and strategies to real-time changes in the operational environment.

Training Environment

The third component of the learning environment involves the nature of the training environment. The nature of the training environment refers to those characteristics under which the training actually occurs, such as (a) the frequency of training opportunities (i.e., how often the training occurs), (b) the length of the training cycle (i.e., how much time transpires between the initiation and completion of the training), (c) the extent to which the training environment simulates the real world (i.e., is the environment realistic and believable), and (d) the location of the training (i.e., is the training conducted in one location or across multiple locations). These factors will impact what can actually be trained in the learning environment. While these characteristics will impact the acquisition and retention of competencies they are rarely discussed in the context of training system design.

Using the frequencies of training as an example, multi-crew IW team training oportunities may be infrequent. Because of the numerous operational requirements placed upon these teams, opportunities for these organizations to train in full complements of the various component commands do not frequently occur. For example, joint and multi-national IW teams may receive opportunities to train together less than one time per year. The limited number of training trials has implications for the retention of skills.

Learning Strategies, Methods, and Tools

Based on the training audience, task requirements, and training environment, a framework which facilitates learning needs to be implemented. The framework needs to provide guidance for what types of learning strategies, methods, and tools can be employed. One framework that has recently demonstrated considerable potential for establishing training in multi-team environments is the Event-Based Approach to Training (EBAT). EBAT ensures that learning occurs by tightly linking critical tasks, learning objectives, exercise design, performance measurement, and feedback. The major EBAT components are diagrammed in Figure 2.

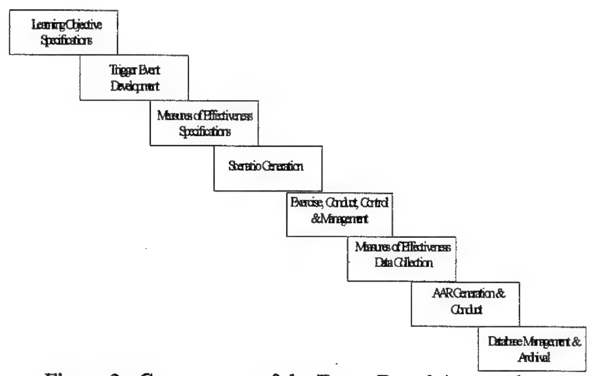


Figure 2. Components of the Event-Based Approach to Training

While other frameworks possess similar characteristics (e.g., Instructional Systems Development; System Approach to Training), EBAT differs from these approaches in two important ways. First, EBAT focuses on the development of scenario-based training, whereas the other approaches to training design were intentionally described as frameworks that can be applied to any training environment. This enables EBAT to address the unique aspects of training in an environment where the scenario is the curriculum. This is contrast to most traditional training settings where a set of instruction (e.g., lessons, lectures, computer-based training) is the curriculum.

Second, EBAT was originally designed for application in team training environments. EBAT methods and tools have been researched and tailored to meet the specific needs found in

team training environments (e.g., performance measurement, feedback). In comparison, the other frameworks have been primarily applied to individual training environments.

Learning environments with components of EBAT have resulted in psychometrically sound measures and improved performance across a variety scenario-based training environments (e.g., Ref 13; Ref 14; Ref 15). The following section will briefly describe each of the components of EBAT. At the end of each description, a brief example of how the EBAT framework can be applied to support multi-crew IW team training will be provided.

EBAT DESCRIPTION AND MULTI-CREW IW TEAM TRAINING EXAMPLE

1) Learning objective specification. EBAT begins with the specification of learning objectives associated with the tasks. Depending on the training audience and task requirements, learning objectives can be associated with a specific task (e.g., demonstrate the ability to perform task XYZ) or general competencies required across a number of tasks (e.g., situation awareness, decision making, resource management, planning, communication). The learning objectives are the foundation of the EBAT framework.

One source of tasks used to guide United Stated Department of Defense (DoD) training design is the Universal Joint Task List (UJTL)(Ref 16). The UJTL provides non-classified listings and descriptions of tasks for a wide range of operations. The following UJTL IW task will be used as the basis for developing the EBAT example.

> Employ Operational Information Warfare. To integrate the use of operations security, military deception, psychological operations, electronic warfare, and physical destruction, mutually supported by intelligence, in order to deny information, influence, degrade, or destroy adversary information, information-based processes, and information systems, and to protect one's own against such actions. As a subset of IW, C2W is an application of IW in military operations that focuses on C2 capabilities.

Based on the information found in the task description, the following four sample IW learning objectives were derived: (a) To protect against enemy actions towards denying access to own information, information-based processes, and information systems. (b) To monitor enemy attempts to influence own information, information-based processes, and information systems. (c) To detect enemy attempts to degrade own force information, information-based processes, and information systems. (d) To communicate enemy attempts to destroy own force information, information-based processes, and information systems. Each of the learning objectives focuses on a specific set of competencies that multi-crew IW teams must possess for effective performance. As an example, the fourth learning objective relates to communication among the team members.

2) Trigger event development. The EBAT framework then requires that "trigger events" be either identified or developed for each learning objective. It should be noted that a trigger event is not an exercise but is instead conditions within an exercise that provides specific opportunities for training audience to practice critical tasks and competencies associated with learning objectives. The events allow the participants to demonstrate their proficiencies and deficiencies for the purpose of performance measurement and feedback. Typically, a number of events are created for each learning objective that (a) vary in difficulty and (b) occur at different points of an exercise.

On the basis of the four learning objectives, sample trigger events were developed (See Table 1). The trigger events will provide specific opportunities for the training audience to demonstrate their ability to conduct defensive multi-crew IW. While the current example listed only two trigger events per learning objective, effective scenario development may require more events to successfully achieve a learning objective. Additionally, the timing and frequency by which trigger events will be introduced into the scenario need to be defined *a priori*.

Learning Objective	Trigger Events
(a) To protect against enemy actions towards denying access to own information, information-based processes, and information systems.	- enemy intrusion on training audience IW systems such that databases can not be accessed - enemy attacks on training audience communication systems
b) To monitor enemy attempts to influence own information, information-based processes, and information systems.	- enemy introduction of data into training audience information sources - enemy modification of training audience databases
c) To detect enemy attempts to degrade own force information, information-based processes, and information systems.	- enemy recon team destroys training audience battlefield sensor system - enemy scrambles training audience radio frequencies
(d) To communicate enemy attempts to destroy own force information, information-based processes, and information systems.	- enemy artillery raids on training audience information system facilities - enemy introduction of computer virus into training audience computer network

Table 1. Sample Learning Objectives and Trigger Events

3) Measures Of Effectiveness (MOE) specification. The EBAT methodology then involves the development of performance measurement strategies and tools required to collect data associated with the trigger events. Depending on the specific characteristics of the learning environment, different measurement strategies and tools may be required. For example, the measurement of competencies that are unique to a single task, given a specific set of conditions (e.g., perform a specific peacekeeping mission in XYZ country that possesses ABC weapons capabilities), are likely to be different from the measurement of competencies that can generalize across a variety of tasks and conditions (e.g., perform strategic planning, situation assessment, and decision making).

A second important characteristic of measurement involves the ability to collect data involving outcomes (e.g., was the right decision made?) and processes (e.g., was the decision made right?) (Ref 12, Ref 13). While measurement of outcomes do provide important information regarding overall performance, measurement of processes is critical for diagnosing specific deficiencies associated with how a given outcome was reached.

The measurement strategies and tools must enable the (a) examination of performance trends during the exercise and (b) development of diagnostic performance feedback. Measuring performance at several events for a specific learning objective enables the development of profiles of how well a team performs on that objective over a range of conditions. Without effective performance measurement and feedback, there is no way of knowing or ensuring--with any degree of certainty--that the training will have its intended effect.

Sample outcome measures related to the overall multi-crew IW task are found in Table 2. While process measures should be developed for each learning objective, Table 3 provides a process measure sample that could be used to collect data related to the fourth learning objective. This particular process measure allows observers to record instances of effective and ineffective communications (i.e., exchanges between multiple crews) in response to enemy attempts to destroy own force information.

Percent	Of attempted adversary penetrations of friendly information systems, successful.
Percent	Of attempted penetrations of adversary information systems successful and apparently not detected.
Percent	Of friendly operations disrupted (because of enemy's interference with friendly information systems).
Percent	Of successful penetrations of adversary info systems detected.
Percent	Of adversary penetrations of friendly info systems, source identified and targeted.

Table 2. Sample Multi-Crew IW Team Outcome Measures

LEARNING OBJECTIVE (D) To communicate enemy attempts to destroy own force information, information-based processes, and information systems. Ground Truth (Scenario Event)	Observer Notes + Phraseology + Brevity + Completeness of Reports + Clarity
08:00	
09:00	
10:00 (Enemy Artillery Raids on Training Audience Information System Facilities)	
11:00	
12:00 (Enemy Introduction of Computer Virus into Training Audience Computer Network)	
13:00	

Table 3. Sample Multi-Crew IW Team Process Measure for Communication

4) Scenario generation. Given the task requirements, learning objectives, trigger events, and performance measures, a scenario is then developed. Scenarios must permit the training audience to interact in realistic situations that will facilitate learning. A variety of constructive, virtual, synthetic, and live resources can be used to develop scenarios. Regardless of the specific resource used, the scenario must be capable of supporting the linkages found among the components of the EBAT framework. A major component of scenario generation involves the development of a master scenario event list (MSEL). Although a MSEL can include information about the hardware/software resources for the scenario, it should also include the specific timing and placement of trigger events within a scenario and their relationship to pre-defined learning objectives (LO) (See Table 4).

LO	Trigger Events	Introduction
A	enemy attacks on training audience communication systems	Day: 1 Time: 08:03
B	- enemy introduction of data into training audience information sources	Day: 1 Time: 09:17
C	- enemy recon team destroys training audience battlefield sensor system	Day: 2 Time 13:13
D	enemy introduction of computer virus into training audience computer network	Day: 2 Time: 20:44

Table 4. Sample Section of Master Scenario Event List

decisions in a manner consistent with doctrine, procedures, and rules of engagement, exercise managers must ensure that the right types of opportunities are presented--in a controlled manner--to meet the intended objectives. Controllers must be capable of modifying a scenario in real-time in response to training audience decisions and performance, for ensuring exercise continuity and realism, and for conducting effective data collection.

6) MOE data collection. As the participants perform during the scenario, measurement data are collected to support feedback. Specifically, when an event occurs, performance related to that event is assessed. Data collection can be conducted using a variety of automated, semi-automated, and observer/trainer-based techniques. While automated techniques are more appropriate for collecting data related to overall outcome scores, humans are better able to collect data related to complex team interactions.

7) After-Action Review (AAR) generation and conduct. Performance is documented, analyzed, and packaged to highlight critical teaching points for subsequent feedback. The systematic linkage continues by tying feedback topics to the performance measures, which in turn are linked to the events and learning objectives. This approach provides structure and control to training and ensures internal consistency throughout an exercise. AARs should include feedback related to both outcome and processes. Figures 3 and 4 depict sample outcome and process summaries that could be used to provide feedback to multi-crew IW teams.

OUTCOME SUMMARY

Percent	**Of attempted adversary penetrations of friendly information systems, successful.**
Percent	**Of attempted penetrations of adversary information systems successful and apparently not detected.**
Percent	**Of friendly operations disrupted (because of enemy's interference with friendly information systems).**
Percent	**Of successful penetrations of adversary info systems detected.**
Percent	**Of adversary penetrations of friendly info systems, source identified and targeted.**

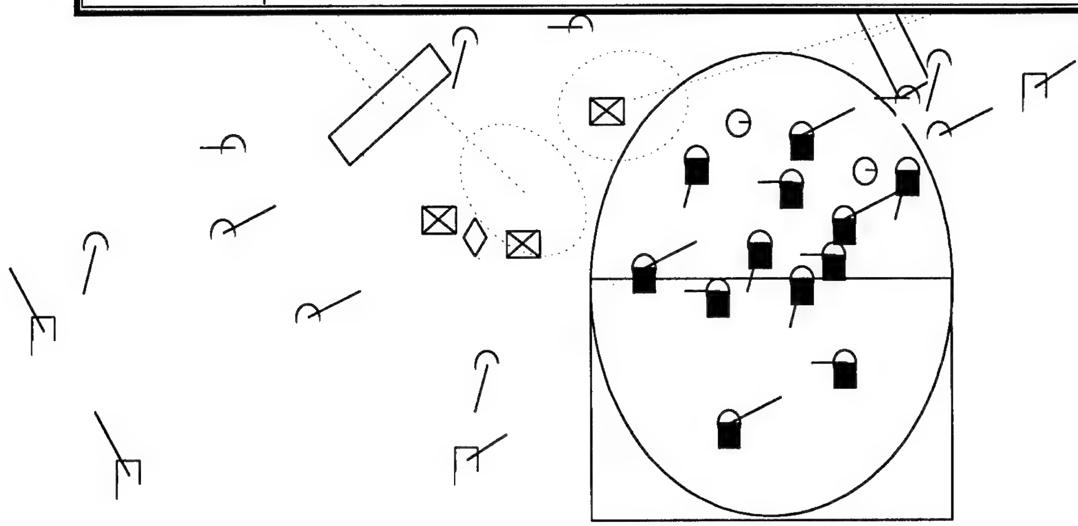


Figure 3. Sample Outcome Summary for Multi-Crew IW Team Performance

LEARNING OBJECTIVE A To protect against enemy actions towards denying access to own information, information-based processes, and information systems.

STRENGTH	
SPECIFIC OBJECTIVE/GOAL :	

LEARNING OBJECTIVE B To monitor enemy attempts to influence own information, information-based processes, and information systems

STRENGTH	
SPECIFIC OBJECTIVE/GOAL	

LEARNING OBJECTIVE C To detect enemy attempts to degrade own force information, information-based processes, and information systems.

STRENGTH	
SPECIFIC OBJECTIVE/GOAL	

LEARNING OBJECTIVE D To communicate enemy attempts to destroy own force information, information-based processes, and information systems.

STRENGTH	
SPECIFIC OBJECTIVE/GOAL	

Figure 4. Sample Process Summary for Multi-Crew IW Team Performance

8) Database management and archival. Following the completion of the exercise, appropriate data are stored and archived in a meaningful manner that supports the development of lessons learned and future exercises. Data collected across exercises can facilitate the development of normative databases. As data accumulate and archives grow, normative patterns will emerge and performance for a given team can be compared against the "norm."

Summary. EBAT provides a framework whereby performance can be traced directly back to specific learning objectives and critical tasks via events and performance measures (See Figure 5).

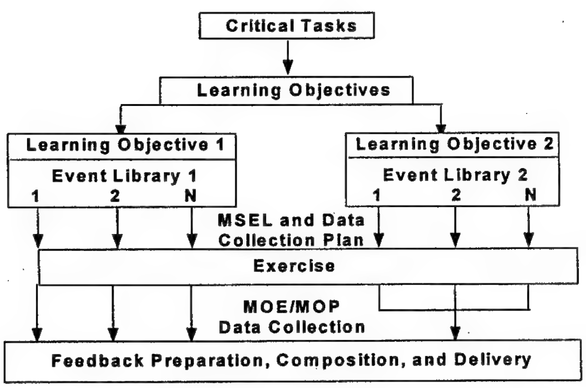


Figure 5. Conceptual Flow between Critical Tasks and the Event-Based Approach for Training

Performance related to a given objective can be assessed and fed back to the training audience. The linkages between each component of EBAT are critical, and therefore must not be viewed as a set of options. If properly implemented, the EBAT framework can be used to establish effective learning environments for team training.

MULTI-CREW IW TEAM TRAINING - CHALLENGES AND CONSIDERATIONS

Clearly, establishing an effective scenario-based training environment for multi-crew IW team is a challenging task. In an effort to meet the challenges, the following considerations are forwarded:

(a) Use a conceptual model of a learning environment as a framework to identify and organize critical factors that will impact learning;

(b) Use EBAT as a set of learning strategies, methods, and tools for establishing a learning environment;

(c) Use learning strategies, methods, and tools that provide systematic linkages among learning objectives, scenario development, performance measurement, and feedback;

(d) Use a multi-faceted approach for performance measurement (e.g., outcome, process, objective, subjective, individual, and team) to support feedback;

(e) Use realistic scenarios that include pre-defined events which provide specific opportunities for the training audience to demonstrate proficiencies and deficiencies related to learning objectives; and

(f) Use scenario control and management techniques that are transparent to the training audience and do not restrict the decisions that can be made by the training audience.

CONCLUSION

The proliferation of IW is likely to become more pronounced in the future as military operations continue to be information intensive. As a result, the ability to conduct effective training for multi-crew IW teams will continue to be important. While continued engineering of training systems is important, technological developments will not be enough.

Because it is enevitable that the nature of IW will change in the future, a continual re-examination of the learning environment characteristics will be required to ensure that appropriate methods, strategies, and tools are developed and implemented to foster learning. Additional work in the development of learning strategies, methods, and tools--such as those offered through EBAT--must be pursued and applied if we expect to maximize training resources.

REFERENCES

1. Alberts, D.S., Papp, D.S., & Tuyahov, A., Berkowitz, B. D. "Historical Impacts of Information Technologies: An Overview". in D.S. Alberts and D.S. Papp (Eds.), "Information Age Anthology: Part One (Vol 1)", Washington DC, USA, National Defense University, 1997, pp. 27-82.

2. Berkowitz, B.D., "Warfare in the Information Age". In D.S. Alberts and D.S. Papp (Eds.), "Information Age Anthology: Part Three (Vol 1)", Washington DC, USA, National Defense University, 1997 (ISBN 1-57906-040-4), pp. 519-544.

3. Alberts, D.S. & Hayes, R.E., "The Realm of Information Dominance: Beyond Information War" in the "First International Symposium on Command and Control Research and Technology", National Defense Institute, June 1995, pp. 560-566.

4. Starr, S.H. & Pace, D.K., Hayes, R.E., "Developing the Intellectual Tools Needed by the Information Warfare Community" in the "First International Symposium on Command and Control Research and Technology", National Defense Institute, June 1995, pp. 577-586.

5. Bell, H. H., "The Engineering of a Training Network" in "The Proceedings of the seventh annual International Training Equipment Conference", Arlington, VA, 1995, pp 365-370.

6. Salas, E. & Cannon-Bowers, J. A., "Methods, Tools, and Strategies for Team Training". In M.A. Quinones & A. Ehrenstein (Eds.), "Training for a Rapidly Changing Workplace: Applications of Psychological Research", Washington, DC, USA, APA Press, 1997, pp. 249-279.

7. Salas, E., Bowers, C. A., & Cannon-Bowers, J. A., "Military Team Research: Ten Years of Progress." Military Psychology, *7*, 1995, pp. 55-75.

8. Neilson, R.E. & Giasson, C.B., "Information – The Ultimate Weapon" in the "First International Symposium on Command and Control Research and Technology", National Defense Institute, June 1995, pp. 545-549.

9. Libicki, M.C., "What is Information Warfare", Washington DC, USA, National Defense University, 1995, pp. 7-8.

10. Cannon-Bowers, J.A., Salas, E, & Converse, S.A., "Shared Mental Models in Expert Decision Making Teams". in N.J. Castellan, Jr. (Ed.), "Current Issues in Individual and Group Decision Making". Hillsdale, NJ: Erlbaum, 1993, pp. 221-246.

11. Urban, J. M., Bowers, C. A., Monday, S. D., & Morgan, B. B., "Workload, Team Structure, and Communication in Team Performance". Military Psychology, 7, 1995, pp. 123-139.

12. Dwyer, D.J., Fowlkes, J.E., Oser, R.L., Salas, E., & Lane, N.E. "Team Performance Measurement in Distributed Environments: The TARGETs Methodology". in M.T. Brannick, E. Salas, & C. Prince (Eds.), "Team Performance Assessment and Measurement: Theory, Methods, and Applications", Mahwah, NJ: Erlbaum, 1997, pp. 137-153.

13. Johnston, J.H., Smith-Jentsch, K.A., & Cannon-Bowers, J.A., "Performance Measurement Tools for Enhancing Team Decision-Making Training". in M.T. Brannick, E. Salas, & C. Prince (Eds.), "Team Performance Assessment and Measurement: Theory, Methods, and Applications", Mahwah, NJ: Erlbaum, 1997, pp. 311-327.

14. Fowlkes, J.E., Lane, N.E., Salas, E., Franz, T., & Oser, R., "Improving the Measurement of Team Performance: The TARGETs Methodology". Military Psychology, *6*, 1994, pp. 47-61.

15. Dwyer, D.J., Oser, R.L., & Fowlkes, J.E., "A Case Study of Distributed Training and Training Performance", in the "Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting", Santa Monica, CA: Human Factors and Ergonomic Society, Oct. 1995, pp. 1316-1320.

16. Chairman of the Joint Chiefs of Staff Manual, 3500.04A, "Univeral Joint Task List - Version 3", Chairman of the Chiefs of Staff, 1996.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO-MP-4 AC/323(HFM)TP/2	ISBN 92-837-1008-8	UNCLASSIFIED/ UNLIMITED

5. Originator	Research and Technology Organization North Atlantic Treaty Organization BP 25, 7 rue Ancelle, F-92201 Neuilly-sur-Seine Cedex, France
6. Title	Collaborative Crew Performance in Complex Operational Systems
7. Presented at/sponsored by	The RTO Human Factors and Medicine Panel Symposium held in Edinburgh, United Kingdom, 20-22 April 1998.

8. Author(s)/Editor(s)	9. Date
Multiple	December 1998

10. Author's/Editor's Address	11. Pages
Multiple	390

12. Distribution Statement

There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.

13. Keywords/Descriptors

Human factors engineering	Computerized simulation
Man machine systems	Cooperation
Situational awareness	Teamwork
Flight crews	Workloads
Group dynamics	Performance evaluation
Decision making	Interpersonal relations
Cognition	Social psychology
Task analysis	Systems engineering

14. Abstract

Contains the proceedings of the first RTO Human Factors and Medical Panel (HFM) Symposium, held in Edinburgh, Scotland, April 20-22, 1998, including the Technical Evaluation Report and Keynote Addresses.

Research and applications in human factors has frequently only considered individual operator interfaces, for limited work domains in well-defined scenarios, as evaluated by unitary measures. As we progress towards the next millennium, complex operations will increasingly require consideration and integration of the collaborative element wherein crew performance becomes a critical factor for success. The goal of this symposium has been to bring together a global perspective on issues and factors that need to be understood when systems design is focused on the crew operating in a complex environment. Hence, the papers contained in these proceedings give the reader a broad, multidisciplinary view of needs, requirements, ongoing research and development projects, and various research agendas that will bring about new technologies, approaches, and measures with regard to collaborative crew performance.

The papers and multiple perspectives contained in these proceedings provide a baseline for understanding many elements of crew performance and in that sense will be valuable for the human factors specialist that must now design for the collaborative element and be concerned with the broad bandwidth of complexities within the operational setting. Additionally, the volume provides information for researchers, scientists, and engineers in many different areas who find themselves immersed in collaborative systems design.

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION
BP 25 • 7 RUE ANCELLE
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • Télex 610 176

DIFFUSION DES PUBLICATIONS RTO NON CLASSIFIEES

L'Organisation pour la recherche et la technologie de l'OTAN (RTO), détient un stock limité de certaines de ses publications récentes, ainsi que de celles de l'ancien AGARD (Groupe consultatif pour la recherche et les réalisations aérospatiales de l'OTAN). Celles-ci pourront éventuellement être obtenues sous forme de copie papier. Pour de plus amples renseignements concernant l'achat de ces ouvrages, adressez-vous par lettre ou par télécopie à l'adresse indiquée ci-dessus. Veuillez ne pas téléphoner.

Des exemplaires supplémentaires peuvent parfois être obtenus auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus sur la liste d'envoi de l'un de ces centres.

Les publications de la RTO et de l'AGARD sont en vente auprès des agences de vente indiquées ci-dessous, sous forme de photocopie ou de microfiche. Certains originaux peuvent également être obtenus auprès de CASI.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE
Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, B-1140 Bruxelles

CANADA
Directeur - Gestion de l'information
(Recherche et développement) - DRDGI 3
Ministère de la Défense nationale
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

ETATS-UNIS
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076

FRANCE
O.N.E.R.A. (Direction)
29, Avenue de la Division Leclerc
92322 Châtillon Cedex

GRECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Aeronautica Militare
Ufficio Stralcio RTO/AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PAYS-BAS
RTO Coordination Office
National Aerospace Laboratory NLR
P.O. Box 90502
1006 BM Amsterdam

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

ROYAUME-UNI
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

TURQUIE
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar - Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076
Etats-Unis

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Document Delivery,
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Les demandes de documents RTO ou AGARD doivent comporter la dénomination "RTO" ou "AGARD" selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
STAR peut être consulté en ligne au localisateur de ressources uniformes (URL) suivant:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR est édité par CASI dans le cadre du programme NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
Etats-Unis

Government Reports Announcements & Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 2216
Etats-Unis
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

Imprimé par le Groupe Communication Canada Inc.
(membre de la Corporation St-Joseph)
45, boul. Sacré-Cœur, Hull (Québec), Canada K1A 0S7

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION

BP 25 • 7 RUE ANCELLE

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE

Telefax 0(1)55.61.22.99 • Telex 610 176

DISTRIBUTION OF UNCLASSIFIED RTO PUBLICATIONS

NATO's Research and Technology Organization (RTO) holds limited quantities of some of its recent publications and those of the former AGARD (Advisory Group for Aerospace Research & Development of NATO), and these may be available for purchase in hard copy form. For more information, write or send a telefax to the address given above. **Please do not telephone.**

Further copies are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO publications, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your organisation) in their distribution.

RTO and AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form. Original copies of some publications may be available from CASI.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, B-1140 Bruxelles

CANADA
Director Research & Development Information Management - DRDIM 3
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92322 Châtillon Cedex

GERMANY
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio Stralcio RTO/AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
See Belgium

NETHERLANDS
RTO Coordination Office
National Aerospace Laboratory, NLR
P.O. Box 90502
1006 BM Amsterdam

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

SPAIN
INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

TURKEY
Milli Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar - Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076
United States

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Document Delivery,
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
STAR is available on-line at the following uniform resource locator:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
United States

Government Reports Announcements & Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 22161
United States
(also available online in the NTIS Bibliographic Database or on CD-ROM)

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

ISBN 92-837-1008-8